# BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.**

## REMARKS

Claims pending in this application include: 283-362, 364-380, 382-398, 400-404, 406-439 and 441-547. Claims amended herein include: 283-294, 360-361, 366, 370, 374-375, 379, 384, 388, 392-393, 397, 402, 408, 411-417, 434, 443-460, 464-475, 479-487, 507, 509-514, 520-524, 526, 528-531 and 546. No claims have been canceled or added by this paper. As required by the new revised format for amendments, a complete set of all the claims in this application has been provided above. Entry of the above amendments is respectfully requested.

### I. Claim Amendments

As indicated above, a number of claims have been amended in this paper. The main purpose of these claim amendments is to define Applicants' claimed invention more clearly. The amendments to the claims are summarized below.

In each of claims 283-294, 360-361, 411-417, 434, 443-460, 464-487 and 546, the term "binding to or" has been deleted with respect to the nucleic acid portion of the non-radioactive signalling entities. Thus, the nucleic acid portion is now defined as being "capable of hybridizing with said bridging entity nucleic acid second portion." This description comports with the annealing practice disclosed in the specification on page 17, last four lines.[1]

In several other claims, dependencies have been changed. These include claims 366, 370, 374-375, 379, 384, 388, 392-393, 397, 402 and 408.

Five claims, 507 and 528-531, have been changed to independent form. These claims were objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

---

**[1] This portion of the specification is also cited in the new matter rejection set forth in the February 17, 2004 Office Action (page 4, penultimate paragraph).**

Claim 509 has also been amended to recite "[a] polynucleotide sequence covalently attached to a *non-nucleotidyl* saccharide having up to 20 saccharide units." Thus, the language in claim 509 has been made clear that the saccharide unit(s) is/are not intrinsic elements of the nucleotide or polynucleotide. Rather, the saccharide units are extra-nucleotidyl elements that are covalently attached to the nucleotide or polynucleotide. A similar change has been effected to claim 522 ("A circular DNA molecule covalently attached to a non-radiolabeled *non-nucleotidyl* signal generating moiety"). Applicants respectfully point out that the term "non-nucleotidyl" is only further defining the nature of the saccharide or saccharide units covalently attached to claimed polynucleotide sequence. The inserted term -- non-nucleotidyl -- is supported by the specification. For example, members of the non-radioactive signalling entities are described on several pages in the specification. For example, beginning on page 96, last paragraph, and continuing through page 97, first paragraph, each of these members are non-nucleotidyl in nature:

> The Sig moiety employed in the make-up of the special nucleotides of this invention could comprise an enzyme or enzymic material, such as alkaline phosphatase, glucose oxidase, horseradish peroxidase or ribonuclease. The Sig moiety could also contain a fluorescing component, such as fluorescein or rhodamine or dansyl. If desired, the Sig moiety could include a magnetic component associated or attached thereto, such as a magnetic oxide or magnetic iron oxide, which would make the nucleotide or polynucleotide containing such a magnetic-containing Sig moiety detectable by magnetic means. The Sig moiety might also include an electron dense component, such as ferritin, so as to be available by observation. The Sig moiety could also include a radiation detecting means. The Sig moiety might also include a hapten component or per se be capable of complexing with an antibody specific thereto. Most usefully, the Sig moiety is a **polysaccharide or oligosaccharide or monosaccharide**, which is capable

> of complexing with or being attached to a sugar or polysaccharide binding protein, such as a lectin, e.g. Concanavilin A. The Sig component or moiety of the special nucleotides in accordance with this invention could also include a chemiluminescent component.
>
> [emphasis added]

Claim 510 has been amended to recite "[a] polynucleotide sequence covalently attached to a *hormonal* receptor." Support for this change is drawn from the specification, page 12, lines 22-23 ("a receptor portion, to be recognized by its hormone; . . .").

Claim 512 has been amended to clarify that the DNA molecule is both *isolated* and that the polynucleotide portion comprises a *repeating low complexity* sequence, the latter comprising *polydCT* and *polydGA*. Claims 513 and 514 both depend from claim 512 and each has been amended above. Amended claim 513 now recites "[t]he DNA molecule of claim 512, *wherein said DNA molecule is derived from a filamentous phage*." Claim 514 now recites "[t]he DNA molecule of claim 513, *wherein said filamentous phage is* a M13 phage or a M13 phage variant."

Similar amendments have been made to other dependent claims, including claims 520, 521 and 523. The former claim now recites "[t]he DNA molecule of claim 516, *wherein said DNA molecule is derived from a filamentous phage*." The latter claim recites "t]he DNA molecule of claim 520, *wherein said filamentous phage* is a M13 phage or a M13 phage variant." Claim 523 in its amended form recites "[t]he DNA molecule of claim 522, *wherein said circular DNA molecule is derived from a filamentous phage*."[2]

With respect to the nature of the filamentous phage recited in various claims that have been amended above, Applicants believe that it is clear from a reading of their disclosure, taken with original claim 79, that the DNA molecule recited in that

---

**[2] See the original specification, page 16, lines 11-20; Example 32; and original claims 31, 32, 44, 45, 70 and 71 (DNA polymer or bridging entity "is *derived* from a filamentous phage").**

original claim is derived from a filamentous phage. This is particularly so because filamentous phages are well known to consist of two portions -- a DNA molecule and a protein coat. The above amendments to the claims make it clear that the DNA molecule now being claimed is derived from a filamentous phage.

Other changes to the claims have been made. Claim 524 now reads that "t]he DNA molecule of claim 522, further [carries] a polynucleotide portion which comprises a *repeating low complexity* sequence selected from the group consisting of poly dGT, poly dAC, poly dCT, poly dAT, poly dGC, poly dGA, poly dG, poly dC, poly dT, poly dA, and *any combination thereof*. The phrase formerly at the end of claim 524, "a sequence or segment of repeating low complexity," has been deleted. In claim 526, the language has been changed to read "[t]he DNA molecule of claim 524, wherein said *repeating low complexity* sequence is *selected from the group consisting of poly dGT, poly dAC, poly dCT, poly dAT, poly dGC, poly dGA, poly dG, poly dC, and any combination thereof*."

Entry of the above amendment to the claims is respectfully requested.

## II. Rejection Based on Written Description

Claims 512-521, 523, 524 and 526 stand rejected for new matter under 35 USC §112, first paragraph. In the February 17, 2004 Office Action (pages 2-4), the Examiner stated:

> It is noted that newly added claims 506-531 have been pointed to by applicants to originally filed claims for written support. Newly added claims 506-531 have been pointed to originally filed claims 72-93 and 96-99, respectivley. Several inconsistencies, however, are present which supports this NEW MATTER rejected due to respective claims as listed above not supplying written basis for the now pending claims.
>
> In newly added claim 512 the last line thereof is not the same as the last line of claim 78 which was pointed to for support. In

particularl original claim 78 is directed to a DNA molecule which is carrying a polynucleotide portion which comprises sequence options as listed therein. The apparently corresponding now pending claim 512 now also is directed to such a polynucleotide but now also includes an option which is an unspecified "sequence" as well as alternatively including a further subdivision of the polynucleotide to include a segment which is of low complexity. This undefined "sequence" option for claim 512 is NEW MATTER as well as the further subdivision of the polynucleotide to contain a "segment" which subdivision was not previously cited in the pointed to claim 78. Claim 524 also contains this NEW MATTER in corresponding inconsistency from original claim 90. Claims which depend directly or indirectly from claims 512 or 524 also contain this NEW MATTER due to their dependency.

Claim 513 also further contains NEW MATTER in that it cites a filamentous phage which "contains" a DNA molecule whereas in contrast originally filed claim 79 discloses that the DNA molecule "is" the filamentous phage. The added "contains" vs. "is" different is NEW MATTER in newly added claim 513. This same NEW MATTER is present in claim 520 which is not consistent with the pointed to originally filed claim 86. Claim 521 also contains this NEW MATTER due to its dependence from claim 520. This NEW MATTER issue is also present in claim 523 which is not consistent with the pointed to originally filed claim 89.

The rejection for new matter is respectfully traversed, or believed to have been obviated by several amendments to the claims above.

**A. <u>Claims 512 and 524</u>**

The Office Action rejects the word "segment" and the second instance of the word "sequence" in claims 512 and 524.

As now amended, both claims no longer include "segment" and the second instance of "sequence."

**B. Claims 513, 514, 520, 521 and 523**

With regard to claims 513, 514, 520, 521 and 523, the Office Action rejects the word "containing" in the phrase "filamentous phage containing the DNA molecule."

As amended above, these claims now recite that the *DNA molecule is derived from a filamentous phage* or that the *filamentous phage is* a M13 phage or a M13 phage variant. It is believed that the language in amended claims 513, 514, 520, 521 and 523 is supported by the original specification, including the originally filed claims directed to the same subject matter. As indicated above, a phage, particularly a filamentous phage, never consists of a DNA molecule and nothing else whatsoever. A phage always includes a DNA molecule plus something else, *e.g.*, a protein component or coat. From a strictly logical standpoint, therefore, a DNA molecule never "is" a phage *per se*. Rather, the DNA molecule is a *component* of the phage, or to be even more technically precise, the DNA molecule is derived from a phage, e.g., filamentous phage.[3] In other words, a phage *comprises* a DNA molecule -- and not the reverse.

In view of the above amendments to the claims, Applicants respectfully request reconsideration and withdrawal of the first new matter rejection.

---

[3] **See footnote 2 above, citing the original specification, page 16, lines 11-20; Example 32; and original claims 31, 32, 44, 45, 70 and 71 (DNA polymer or bridging entity "is *derived* from a filamentous phage").**

**III. Rejection Based on Indefiniteness and Written Description**

Claims 283-362, 364-380, 382-398, 400-404, 406-439, 441-505 and 532-547 stand rejected under 35 USC §112, first and second paragraphs. In the Office Action (pages 4-5), the Examiner stated:

> The claimed compositions, kits, and methods, cite a signaling entity nucleic acid portion per se therein which is capable of either binding or hybridizing with the bridging entity nucleic acid second portion. See, for example, claim 283, lines 7-8, specifically. These binding "or" hybridizing options are not commensurate in scope with the interaction cited for said signaling entity with the bridging second portion which only cites annealing and not generic binding practice. This is directly cited in the specification on page 17, last 4 lines, wherein this annealing practice is cited but not generic binding between the above two cited entities. This limited citation to annealing is also present in originally filed instant claim 1.
>
> The above discussed limitation to annealing as originally filed, which is reasonably interpreted as functionally equivalent to hybridizing, whereas, in contrast, the present claims are broader to include generic binding also supports the below NEW MATTER rejection of the presently pending instant claims.
>
> Claims 283-362, 364-380, 382-398, 400-404, 406-439, 441-505, and 532-547 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. The NEW MATTER has been described in the immediately preceding paragraph.

The currently amended claims no longer recite the word "binding" in the context of the association of the signalling entity's nucleic acid with the bridging entity's nucleic acid second portion. Thus, the rejection of claims 283-362, 364-380, 382-398, 400-404, 406-439, 441-505 and 532-547 has been obviated by the amendments above.

In light of the above claim amendments, Applicants respectfully request reconsideration and withdrawal of the rejection for indefiniteness and new matter.

## IV. Prior Art -- First Anticipation Rejection -- Maniatis et al.

Claims 509, 510, 512-517 and 520-527 stand rejected under 35 U.S.C. §102(b) for anticipatbn by MOLECULAR CLONING [Maniatis et al. (1982)]. In the Office Action (pages 5-6), the Examiner stated:

> Pages 51-54 of Maniatis et al. disclose various filamentous phages which contain DNA molecules as well as both single and double stranded nucleic acid forms thereof which anticipates the above instant claims. Sequences therein are disclosed on page 53 which specifically read on sequences as cited in instant claim 512, for example. Such polynucleotides may also be interpreted reasonably as functional segments therein which also comprise a polysaccharide as required in instant claim 509. It is well known that each nucleotide unit is a saccharide unit as in said claim 509. Page 53 of Maniatis et al. also discloses segments of the M13 DNA which are receptors for restriction endonuclease binding sites as also required in instant claim 510. Page 52 of the reference cites the presence of a covalently attached polylinker segment which produces a detectable signal via insertions therein thus anticipating instant claim 522 etc.

### A. Claim 509

Amended claim 509 now recites a polynucleotide sequence covalently attached to "a *non-nucleotidyl* saccharide" having up to 20 saccharide units. It is

believed that the foregoing recitation regarding the saccharide being *non-nucleotidyl* materially distinguishes claim 509 from Maniatis et al.

**B. Claim 510**

Amended claim 510 now recites a polynucleotide covalently attached to a *hormonal* receptor. The further definition of the claimed receptor being *hormonal* in nature is believed to materially distinguish Applicants' invention from Maniatis et al.

**C. Claims 512-517 and 520-527**

Claims 512-517 and 520-527 are directed to a DNA molecule that includes a repetitive low complexity sequence. The Office Action asserts that claims 512-517 and 520-527 are anticipated by sequences disclosed on page 53 of Maniatis et al.

The basis for this rejection is somewhat unclear. Maniatis et al. disclose a M13 vector with a polylinker that contains restriction sites. Maniatis et al. do not disclose that the polylinker includes a repetitive low complexity sequence. Restriction sites often include palindromic sequences, but persons in the art would not consider a palindromic sequence to be a repetitive low complexity sequence as set forth in amended claims 512-517 and 520-527. Moreover, claims 512, 524 and 526 no longer recite a generic "sequence or segment of repeating low complexity." They now only recite specific repetitive low complexity sequences that are not disclosed in Maniatis et al. As such, Maniatis et al. do not disclose a polynucleotide with any of the particular low complexity sequences now recited in these claims.

Applicants further respectfully point out that currently amended claim 512 explicitly recites that the DNA molecule is "isolated." When the original claims were filed, the Patent Office did not typically require recitation of the term "isolated" in cases where it was obvious that the limitation was implicitly present in the claim. In view of a growing tendency by the Patent Office and the courts to read claims "super literally," Applicant believes that the intended meaning of the original claim is more accurately expressed today by explicitly reciting that the DNA

molecule is *isolated*. Support for such isolated DNA is implicit throughout the original claims and specification.

**D. Claim 522**

As amended above, claim 522 is directed to a circular DNA molecule covalently attached to a non-radiolabeled *non-nucleotidyl* signal generating moiety. The Office Action says that the polylinker of Maniatis et al "produces a detectable signal via insertions therein thus anticipating claims 522 etc." The inclusion of a *non-nucleotidyl* signal generating moiety in claim 522 represents a material element altogether lacking in Maniatis et al.

In view of the foregoing remarks and the above amendments to the claims, Applicants respectfully request reconsideration and withdrawal of the first anticipation rejection.

## V. Prior Art -- Second Anticipation Rejection -- Langer et al.

Claims 506, 509 and 510 stand rejected under 35 USC §102(b) for anticipation by Langer et al. [PNAS 78(11) 6633 (1981)]. In the Office Action (page 6), the Examiner stated:

> Langer et al. discloses polynucleotides of DNA or RNA with biotin covalently attached which covalent attachment also results in antibody attachment which anticipates the above listed instant claims. Such an antibody is also a receptor for biotin binding as required in certain instant claims such as claim 510. See the entire document and especially page 6635, first column, second full paragraph.

The second anticipation rejection is respectfully traversed.

Applicants respectfully point out that the cited Langer et al. document does not disclose a polynucleotide sequence covalently attached to an antibody. Instead, Langer et al. disclose a biotin molecule covalently attached to a DNA molecule and an antibody that is noncovalently attached to the previously attached biotin molecule. To state it in another way, in Langer et al, biotin is *not* covalently

attached to an antibody. Biotin binds *non-covalently* to avidin.[4] Biotin and avidin bind to each other via ionic, polar and/or hydrogen bonds.[5] Because a non-covalent bond is always interposed between biotin and any antibody attached to avidin, the indirect attachment of biotin to the antibody on avidin cannot be characterized as covalent. As such, the antibody in Langer et al. cannot reasonably be construed to be covalently attached to a polynucleotide sequence as set forth in claim 506. Thus, Langer et al. lack a material element recited in claim 506, namely, an antibody that is covalently attached to a polynucleotide sequence.

In view of the foregoing remarks and the clear lack of identity in material elements between the subject matter recited in claim 506 and Langer's disclosure, Applicants respectfully request reconsideration and withdrawal of the second anticipation rejection.

## VI. Claim Objections

The Office Action rejects claims 507 and 528-531 for depending from rejected claims. As indicated in the opening remarks of this paper, these five claims have been rewritten in independent form, and thus, should be allowable.

## VII. Dependencies from Cancelled Claims

Applicant has identified various dependent claims that, prior to this Amendment, depended from previously cancelled claims. These errors have been fixed by the amendments to the claims above, and include the following:

---

[4] *See* Langer et al, US Patent No. 4,711,955, col. 1, lines 44-47, col. 18, lines 19-21, previously cited of record in this application; see also Laitinen et al., "Chicken avidin-related proteins show altered biotin-binding and physico-chemical properties as compared with avidin," Biochem J. 363:609-17 (2002); see in particular the abstract on page 609. A copy of Laitinen et al. is attached as Exhibit 1.

[5] *See* Weber et al., "Structural Origins of High-Affinity Biotin Binding to Streptavidin," Science 243:85-88 (1989). A copy of Weber et al. is attached as Exhibit 2.

Currently amended claims 366, 370, 374, 375 and 379 now depend from claim 443 rather than from previously cancelled claim 363.

Currently amended claims 384, 388, 392, 393 and 397 now depend from claim 444 rather than from previously cancelled claim 381.

Currently amended claim 402 now depends from claim 445 rather than from previously cancelled claim 399.

Currently amended claim 408 now depends from claim 446 rather than from previously cancelled claim 405.

## VIII. Submission of Information Disclosure Statement

Applicants and their attorney(s) are presently reviewing materials and files for the purpose of submitting art-related documents in an Information Disclosure Statement. As soon as any such documents have been located, their undersigned attorney will promptly submit them in an IDS.

Favorable action is requested.

**SUMMARY**

Amended in this paper are claims 283-294, 360-361, 366, 370, 374, 375, 379, 384, 388, 392, 393, 397, 402, 408, 411-417, 434, 443-460, 464-475, 479-487, 507, 512-514, 520-524, 526 and 528-531. No new claims are added. No fee is believed due in connection with the filing of this Amendment. If any fee is due, however, The Patent and Trademark Office is hereby authorized to charge the amount of any such fee to Deposit Account No. 05-1135, or to credit any overpayment thereto.

If the Examiner has any questions, he is invited to contact the undersigned attorneys.

Respectfully Submitted

Dated: APRIL 29 2004 By: _____

Samson Vermont, Reg. No. 42,202
Robert M. Schulman, Reg. No. 31,196
Ronald C. Fedus, Reg. No. 32,567

Hunton & Williams
1900 K Street, N.W.
Suite 1200
Washington, D.C. 20006-1109
Telephone: (202) 955-1500
Facsimile: (202) 778-2201

Enzo Biochem, Inc.
527 Madison Avenue, 9th Floor
New York, NY 10022-4304
Telephone: (212) 583-0100
Facsimile: (212) 583-0150

Biochem. J. (2002) 363, 609–617 (Printed in Great Britain)



# Chicken avidin-related proteins show altered biotin-binding and physico-chemical properties as compared with avidin

Olli H. LAITINEN*[1], Vesa P. HYTÖNEN*, Mervi K. AHLROTH*, Olli T. PENTIKÄINEN†, Ciara GALLAGHER*, Henri R. NORDLUND*, Vladimir OVOD‡, Ari T. MARTTILA*, Eevaleena PORKKA*, Sanna HEINO*[2], Mark S. JOHNSON†, Kari J. AIRENNE*[1] and Markku S. KULOMAA*[3]

*Department of Biological and Environmental Science, University of Jyväskylä, P.O. Box 35, FIN-40351 Jyväskylä, Finland, †Department of Biochemistry and Pharmacy, Åbo Akademi University, P.O. Box 66, FIN-20521 Turku, Finland, and ‡Institute of Medical Technology, University of Tampere, FIN-33014 Tampere, Finland

Chicken avidin and bacterial streptavidin are proteins familiar from their use in various (strept)avidin–biotin technological applications. Avidin binds the vitamin biotin with the highest affinity known for non-covalent interactions found in nature. The gene encoding avidin (*AVD*) has homologues in chicken, named avidin-related genes (*AVR*s). In the present study we used the *AVR* genes to produce recombinant AVR proteins (AVRs 1, 2, 3, 4/5, 6 and 7) in insect cell cultures and characterized their biotin-binding affinity and biochemical properties. Amino acid sequence analysis and molecular modelling were also used to predict and explain the properties of the AVRs. We found that the AVR proteins are very similar to avidin, both structurally and functionally. Despite the numerous amino acid substitutions in the subunit interface regions, the AVRs form extremely stable tetramers similar to those of avidin. Differences were found in some physico-chemical properties of the AVRs as compared with avidin, including lowered pI, increased glycosylation and, most notably, reversible biotin binding for two AVRs (AVR1 and AVR2). Molecular modelling showed how the replacement $Lys^{111}$→isoleucine in AVR2 alters the shape of the biotin-binding pocket and thus results in reversible binding. Both modelling and biochemical analyses showed that disulphide bonds can form and link monomers in AVR4/5, a property not found in avidin. These, together with the other properties of the AVRs described in the present paper, may offer advantages over avidin and streptavidin, making the AVRs applicable for improved avidin–biotin technological applications.

Key words: avidin–biotin technology, biotin-binding protein, molecular evolution, molecular modelling, structure–function relationship.

## INTRODUCTION

Chicken avidin and bacterial streptavidin are known for their extraordinarily high affinity for a water-soluble vitamin, biotin [1]. Owing to their high affinity and specificity for biotin, avidin and streptavidin have been utilized in numerous applications in life sciences, including purification, labelling and targeting of various materials. The methodology has been collectively referred to as (strept)avidin–biotin technology [1,2].

Avidin and streptavidin are tetrameric proteins consisting of four identical subunits. Their three-dimensional (3-D) structures have been solved by X-ray crystallography, and their tertiary and quaternary structures show astonishing similarity, despite relatively low amino acid sequence identity [3–6]. Most of the essential biotin-binding residues are conserved, and the affinity for biotin is nearly identical in both avidin and streptavidin. In addition to high biotin-binding capacity, avidin and streptavidin tetramers show remarkable stability at high temperatures, as well as under strongly denaturing conditions. The stability increases even more upon biotin binding [7–9].

There are, however, prominent differences in some biochemical properties of avidin and streptavidin. Streptavidin is not glycosylated and lacks cysteine residues capable of forming disulphide bridges, whereas avidin has a carbohydrate side chain and one intramolecular disulphide bond. Furthermore, the pI of streptavidin is acidic (pI ≈ 6) in contrast with basic in avidin (pI ≈ 10.5) [10]. From the higher-order structural point of view, the interfaces between the subunits are built somewhat differently in these two proteins. The striking similarity in some properties with simultaneous disparity in others, coupled with a long evolutionary distance between them, make avidin and streptavidin an ideal model system for studying the evolution of ligand-binding proteins.

The avidin gene (*AVD*) in chicken has homologues, called avidin-related genes (*AVR*s) [11–13]. The number of *AVR* genes seems to vary between individuals [14], but seven different genes, *AVR1*–*AVR7*, have been cloned and sequenced. Two of the genes, *AVR4* and *AVR5*, are 100% identical in their coding sequence, exhibiting only a single nucleotide difference in their 5′-flanking region, whereas the others are 94–99% identical to each other. The identity between *AVD* and the different *AVR*s ranges from 91 to 95% [12,13]. mRNAs for *AVR1*, *AVR2* and *AVR3* have been detected in chicken under inflammatory conditions ([15]; P. Lappalainen, T. Kunnas, E.-L. Punnonen and M. S. Kulomaa, unpublished work), but it is not known whether they are expressed as proteins. The putative avidin-

Abbreviations used: *AVR*, avidin-related gene; *AVD*, avidin gene; AVR, avidin-related protein; 3-D, three-dimensional; RT-PCR, reverse-transcription PCR; Endo $H_f$, recombinant endoglycosidase H fused ($_f$) to maltose-binding protein; PNGase F, peptide:N-glycosidase F; IgG–AP, IgG–alkaline phosphatase conjugate; L3 (etc.), loops designated by their numeral orders; $\beta$3 (etc.), $\beta$-sheets designated by their numeral orders; $K_d$, dissociation constant; $k_{diss}$, dissociation rate constant; $k_{ass}$, association rate constant; CPK, Corey–Pauling–Koltun.

[1] Present address: A. I. Virtanen Institute, Department of Molecular Medicine, University of Kuopio, P.O. Box 1627, Kuopio, Finland.

[2] Present address: National Public Health Institute, Department of Molecular Medicine, University of Helsinki, P.O. Box 104, FIN-00251 Helsinki, Finland.

[3] To whom correspondence should be addressed (e-mail markku.kulomaa@csc.fi).

related proteins (AVRs), as deduced from their nucleotide sequences, are 74–81% identical with avidin and 85–100% identical with each other.

In our previous studies on the structure–function relationship of avidin, we employed a variety of different approaches to examine the extent to which amino acid residues of avidin can be mutated without affecting biotin binding and stability. One successful strategy has been to use an evolutionary approach. For example, we have engineered avidin according to the putative AVR sequences to lower the pI of avidin down to 4.7, as well as to remove its N-glycosylation site [16,17]. Also, the avidin-like domain of sea-urchin fibropellins was used as a model to dissociate the avidin (and streptavidin) tetramer into stable dimers [18]. A structural approach was used in designing mutations to produce a monomeric avidin [19].

The successful use of AVR sequences as models to improve the properties of avidin encouraged us to perform a detailed analysis of the sequence variability in terms of modelled 3-D structures for the putative AVR proteins. The AVRs seem to be well conserved, with most of the biotin-binding residues preserved. However, they exhibit a number of amino acid replacements that may affect their physico-chemical characteristics as compared with avidin. The structural and functional properties of the AVR proteins are of considerable interest to new avidin–biotin technology.

## EXPERIMENTAL

### Sequence comparison and molecular modelling

The EMBL database accession numbers for the *AVR* genes used to deduce the corresponding amino acid sequences are as follows: *AVR1*, Z21611; *AVR2*, Z21554; *AVR3*, Z21612; *AVR4*, Z22883, *AVR6*, AJ237658; *AVR7*, AJ237659. *AVR4* represents also the identical *AVR5*, and we therefore use here the term *AVR4/5*. The *AVR* cDNAs were translated with the GCG software package program Map (Genetic Computer Group, Madison, WI, U.S.A.). The theoretical pI for the AVRs and avidin was determined by using the GCG-package program Peptidesort. The comparison of the sequences of avidin and AVR1–AVR7 was performed using Malign [20,21] in the BODIL Modelling Environment (J. Lehtonen, V.-V. Rantanen, D.-J. Still and M. S. Johnson, unpublished work). Model structures of AVR1–AVR7 were made using the rapid homology-modelling program HOMODGE in BODIL on the basis of the X-ray structure of chicken avidin in complex with biotin {Research Collaboratory for Structural Bioinformatics PDB (Protein Data Bank) code: 1avd; [6]}. HOMODGE reduces errors in models of closely related proteins by using, as much as possible, atomic coordinates from the known structure to build the conserved main-chain and side-chain structures. Where the side chains differ, HOMODGE uses a rotamer library to select the side-chain conformation for optimal contacts.

### Production and purification of recombinant AVR proteins

The *AVR1–4/5* genes [11,12] were cloned into the *Eco*RI–*Hin*dIII sites of pGEM4 and were *in vitro* transcribed and spliced using Ribomax and RNA Splicing System kits (Promega) according to the manufacturer's instructions. The cDNAs were then produced by reverse-transcription (RT)-PCR and further amplified by PCR using the oligonucleotide primers: AK33 (5′-CTGCTAGATCTATGGTGCACGCAACCTCCCC-3′) and AK44 (5′-GTTGCAAGCTTTGCGGGGCCATCCT-3′) containing *Bgl*II and *Hin*dIII restriction sites respectively. After cutting with *Bgl*II and *Hin*dIII, the AVR1–4/5 cDNAs were cloned into *Bam*HI and *Hin*dIII sites of pFASTBAC. For *AVR6* and *AVR7*, cDNAs were produced by subcloning the corresponding genes [13] into pDsRed1 (ClonTech) where the red-fluorescent-protein-encoding region had been removed, and by subsequent transfection of the constructs into NIH/3T3 cells. Total RNA was extracted from the cells using the SV Total RNA Isolation System (Promega). The AVR6 and AVR7 cDNAs were produced by RT-PCR (RobusT RT-PCR Kit; Finnzymes, Espoo, Finland) using the oligonucleotide primers AK33 and AK44 to produce *Bgl*II and *Hin*dIII restriction sites to the 5′- and 3′-ends of the cDNAs respectively. The synthesized cDNAs were cloned into *Bam*HI/*Hin*dIII-digested pFASTBAC1, the cloning vector for the Bac-To-Bac Baculovirus Expression System (Gibco BRL, Life Technologies, Gaithersburg, MD, U.S.A.). The nucleotide sequences of the cDNAs were confirmed by sequencing. The virus vectors for producing the AVR proteins were constructed and amplified according to the Bac-To-Bac system instructions. Recombinant AVR proteins were produced in *Spodoptera frugiperda* (Sf9) insect cells as previously reported [22]. Proteins were purified from the cells using affinity chromatography on a 2-iminobiotin (AVR1, AVR3, AVR4/5–7) and/or biotin–agarose column (AVR1 and AVR2) as previously described [18]. AVR2 was eluted from biotin–agarose with 1 M acetic acid and the eluted fractions were immediately neutralized with NaOH. The elution of AVR1 was achieved using 1 M HCl, followed by neutralization with Tris (1 g/ml).

### Biotin-binding analyses

The biotin-binding characteristics of the AVR proteins were studied as previously described using the IAsys optical biosensor (Thermo Labsystems, Helsinki, Finland) [16,18]. Affinities were measured for 2-iminobiotin or biotin at 20 °C using a stirring power of 100%. Briefly, the protein in biotin-free buffer was allowed to bind to the biotin-coated surface of the IAsys cuvette. As a negative control, the protein was saturated with an excess of biotin prior to addition into the cuvette. Excess of biotin was also maintained during analysis. Under these conditions, binding to the biotin surface was regarded as unspecific and used as a measure of baseline variation. Steady-state equilibrium in the experimental cuvette was considered to be attained after no additional binding was detected as compared with the negative control.

In order to study the reversibility of biotin binding, avidin and the AVRs were allowed to bind to a biotin-coated cuvette. After measuring the maximal binding ($A$) as described above, biotin-saturated (0.17 mg/ml) buffer was injected into the cuvette. The amount of protein remaining bound ($B$) was measured after no further dissociation was seen as compared with the negative control. The dissociation was measured for at least 20 min. The percentage reversibility was calculated as follows:

$$\textit{Reversibility}\ (\%) = 100 \times (A - B)/A$$

$A - B$ is the amount of liberated proteins after addition of biotin.

### Structural analyses

The heat-stability of the AVRs was studied by using an SDS/PAGE-based method [23]. In this method, each protein sample was divided into two fractions, one of which was kept biotin-free and the other was supplemented with an excess of biotin. The samples were then mixed with denaturing SDS/PAGE sample buffer (containing $\beta$-mercaptoethanol) and incubated at a given temperature for 20 min. After heat treatment, the samples were subjected to SDS/PAGE and visualized by

**Table 1 Amino acid substitutions in the interface regions between AVRs and avidin**

Amino acids, which in avidin participate in intersubunit interactions but are substituted in AVRs. Amino acid substitutions: H50L (etc.), $His^{50} \rightarrow$ leucine (etc.). Abbreviation: h-phobic, hydrophobic.

| Amino acid ... Change and type of change ... | H50L Basic/ h-phobic | N54H Polar/ basic | N57A Polar/ neutral | R59A Basic/ neutral | N69H Polar/basic | K71N Basic/polar | T80V Neutral/ h-phobic | M96K h-phobic basic | S101L polar/ h-phobic | I117N/Y h-phobic/ polar |
|---|---|---|---|---|---|---|---|---|---|---|
| Occurrence in AVRs | All | All | All | All | All | All | All but 4/5 | All but 4/5 | All but 4/5 | AVR2, AVR3, N; the rest, Y |
| Side/main chain | M*/O† | S‡ | S | S | S | S | M/O | S | M/O | S |
| Hydrogen bond/VdW§ | 2¶ H | 7 H | 1 H | 2 H | 3 H | 1 H | 2 H | VdW | 1 H | VdW |
| Interface contacts | 1/4 | 1/4 | 1/4 | 1/4 | 1/4 | 1/4 | 1/4 | 1/3 and 1/4 | 1/4 | 1/3 |

* Main-chain contact.
† Oxygen atom of main chain.
‡ Side-chain contact.
§ Van der Waals contacts.
¶ Number of hydrogen bonds.

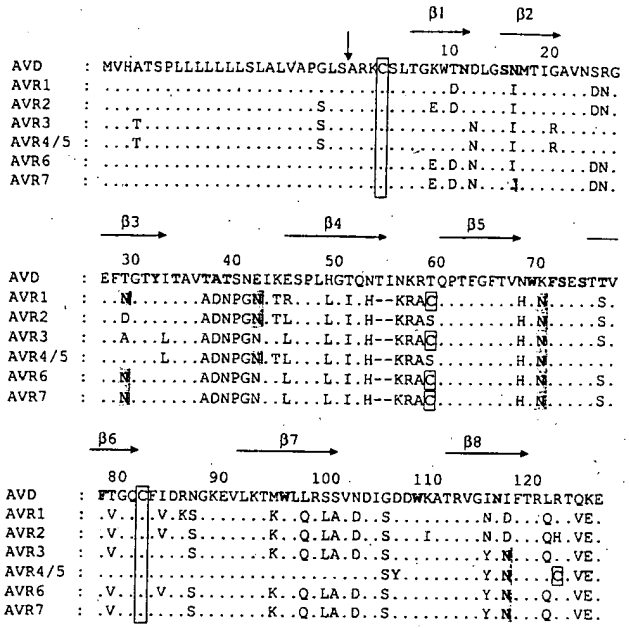


**Figure 1 Multiple sequence alignment of avidin (AVD) and the AVRs**

Dots indicate identical amino acids in AVRs as compared with avidin, and the two amino acid deletion in AVRs is indicated by dashes. Horizontal arrows designate the $\beta$-sheets of avidin, and the vertical arrow indicates the cleavage site of the signal peptide in avidin. Biotin-binding residues are in **bold** in the avidin sequence, and the N-glycosylation site of avidin as well as the potential N-glycosylation sites of the AVRs are highlighted with grey. Cysteine residues are boxed.

staining with Coomassie Brilliant Blue. The relative proportions of tetrameric and monomeric forms of the AVR proteins (with avidin as the control) were detected. A similar assay, with $\beta$-mercaptoethanol omitted from the sample buffer, was performed to examine the presence of intersubunit disulphide bridges. Sensitivity to proteinase K was studied in both the absence and presence of biotin as described in [18]. The pI of each AVR was determined by isoelectric focusing as previously reported [16]. The glycosylation patterns of the AVR proteins were studied by treating the proteins with recombinant endoglycosidase H fused ($_r$) to maltose-binding protein (Endo $H_r$) and PNGase F (peptide: N-glycosidase F) as described in the manufacturer's (New England Biolabs) instructions.

## Immunological analyses

In order to investigate whether the immunological properties of AVR proteins differ from those of avidin, an indirect ELISA analysis was utilized [24]. The wells of a 96-well plate were coated with avidin or AVR proteins (1 $\mu$g/ml) in 50 mM sodium carbonate buffer at 37 °C for 2 h, followed by washing in PBS/Tween and blocking with 1 % BSA in PBS. ELISAs were performed using two monoclonal anti-avidin antibodies produced in one of our institutions (FIT Biotech), and a polyclonal rabbit anti-avidin antibody (produced at the Laboratory Animal Center, University of Oulu, Oulu, Finland) as primary antibodies. Goat Anti-Mouse IgG–alkaline phosphatase (Bio-Rad) or goat anti-rabbit IgG–alkaline phosphatase (Bio-Rad) was used as secondary antibodies, respectively. The used signal molecule was *p*-nitrophenyl phosphate (1 mg/ml) (Sigma). $A_{405}$ values were measured with an automated ELISA reader.

## RESULTS

### Sequence comparison and molecular modelling

The putative AVRs were compared with avidin to identify differences that might affect their biotin binding, structural or other biochemical properties. The biotin-binding residues are well conserved in all AVR proteins, showing only three amino acid changes (out of the 16 possible) (Figure 1). Three sequential residues involved in biotin binding (38–40; Thr-Ala-Thr in avidin) are replaced by Ala-Asp-Asn in all AVR proteins. With the exception of AVR7, the AVR proteins lack the glycosylation site ($Asn^{17}$) present in avidin. Instead, they all exhibit other Asn-Xaa-Ser/Thr sequences, which could be glycosylated during protein maturation along the secretion route. There is one common Asn-Xaa-Ser site in the L5 loop of all the AVR proteins. Additional Asn-Xaa-Ser/Thr sequences are distributed as follows: two in AVR1 (in strand $\beta$3 and in L3), one in AVR2 (L3), one in AVR3 ($\beta$8), two in AVR4/5 (L3 and $\beta$8), and two in AVR6 and AVR7 ($\beta$3 and $\beta$8). The potential N-glycosylation sites of the AVRs are located on the surface of the tetramers and, therefore, they are expected to be glycosylated.

The theoretical pI values for AVR3 and AVR4/5 are basic, similar to the pI of avidin (avidin, 10.4; AVR3, 10.2; and AVR4/5, 10.0). On the other hand, AVR1, AVR6 and AVR7 are neutral (pI 7.3), and AVR2 is calculated to be acidic with a pI of 4.7. All AVR proteins, except for AVR2, have three cysteine residues (Figure 1). In AVR1, AVR3, AVR6 and AVR7, the third 'extra' cysteine replaces $Thr^{60}$ of avidin located at the

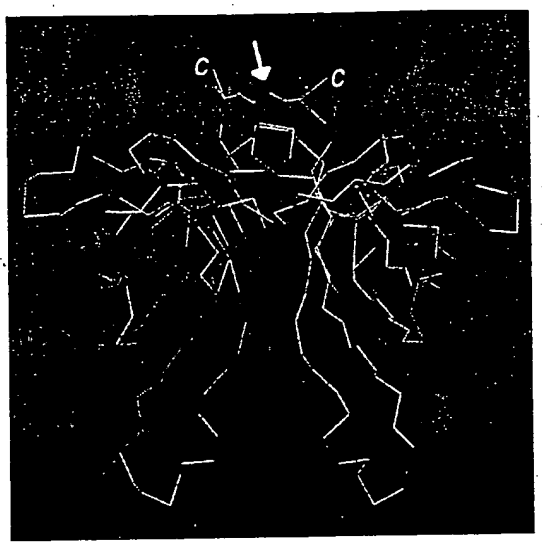


**Figure 3 Model structure for two monomers of AVR4/5**

Compared with the avidin structure, AVR4/5 has an additional cysteine residue in each monomer located at the C-terminus (arrow). This dimer model, based on the avidin structure, would place these two cysteine residues in an ideal location to form a disulphide bond between the two monomers, but this extra disulphide bond would not prevent the formation of a tetramer similar to that seen for avidin.

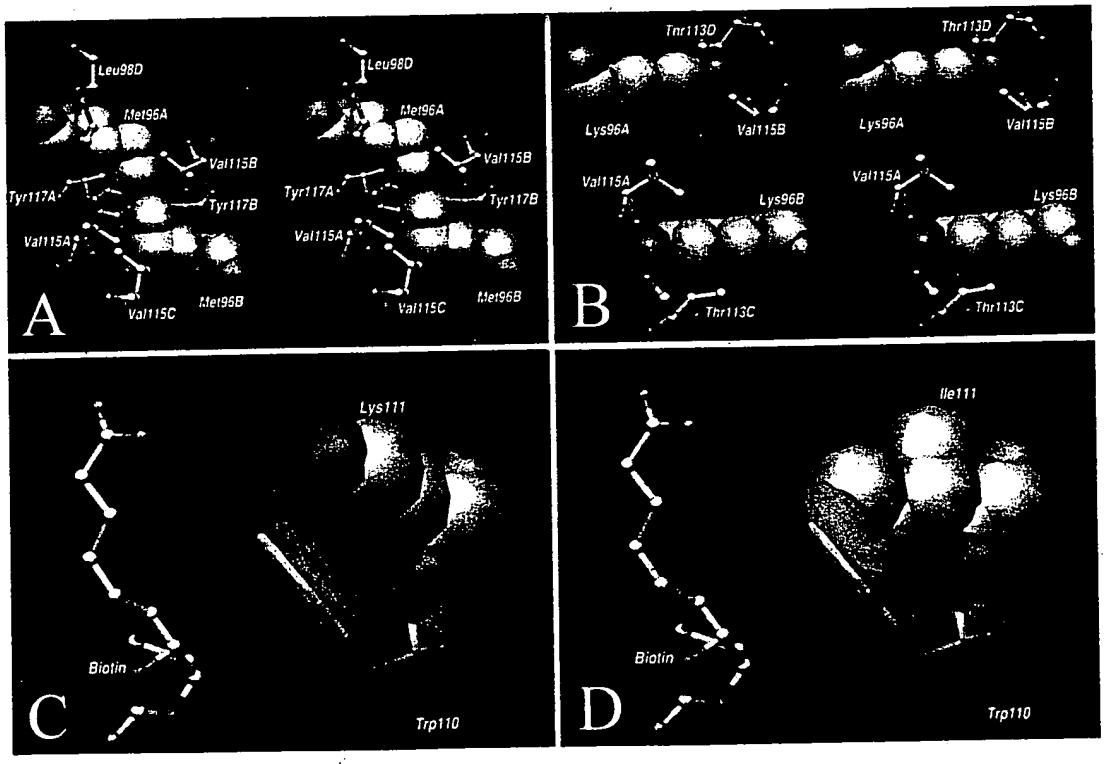


**Figure 2 Re-organization of the subunit interface where sequence differences at position 96 are located (in stereo) and the effect of sequence differences at position 111 on the orientation of the side chain of $Trp^{110}$**

The amino acid at position 96 is shown as a CPK (Corey–Pauling–Koltun) model and the surrounding amino acids are shown as ball-and-stick representations. (**A**) $Met^{96}$ (avidin and AVR4/5); (**B**) $Lys^{96}$ (AVR1–3, AVR6 and AVR7); (**C**) $Lys^{111}$ (avidin, AVR1, AVR3–7); (**D**) $Ile^{111}$ (AVR2) as CPK models. $Trp^{110}$ is drawn as sticks along with the solvent-accessible (transparent) surface; ball-and-stick models represent the ligand biotin. Note in (**D**) the severe overlap between $Ile^{111}$ and $Trp^{110}$ that would force the reorientation of the tryptophan side chain, leading to weaker interactions with biotin and reduced binding affinity.

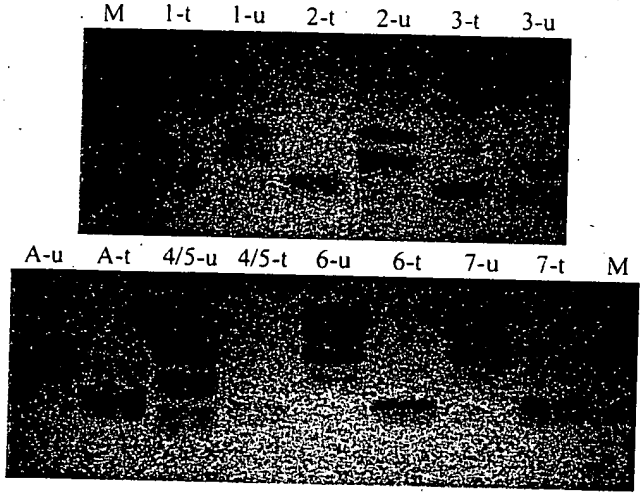


**Figure 4 Glycosylation of avidin and AVRs in baculovirus-infected insect cells**

Samples were either treated or non-treated with Endo $H_f$ glycosidase followed by SDS/15%-(w/v)-PAGE and staining with Coomassie Brilliant Blue. Avidin is denoted by 'A' and the different AVRs by their corresponding numbers. 'M' indicates the low-molecular-mass markers (31, 21.5 and 14.4 kDa; Bio-Rad). Abbreviations: -u, not treated with enzyme; -t, treated with enzyme.

beginning of $\beta$5. In AVR4/5, the 'extra'cysteine residue is close to the C-terminus of the protein where $Arg^{124}$ is found in avidin.

The interface regions between different subunits show a variable number of amino acid substitutions between the AVR proteins and avidin (see [5,6] for the interface structures of avidin). The interface between subunits 1 and 2 (as well as between 3 and 4) is perfectly conserved in all of the AVRs. In contrast, two out of three interface residues between subunits 1 and 3 (2 and 4) show changes in all of the AVR proteins, except for AVR4/5, which shows only one substitution ($Ile^{117}$→tyrosine). Interestingly, the tightly interacting 1–4 (2–3) subunit interface shows several amino acid substitutions in all of the AVRs. In avidin, a total of 22 residues confer intersubunit contacts within this interface. The AVR proteins have nine substitutions in this region, seven of which are found in all AVRs and two in all AVRs but AVR4/5 (Table 1).

Most of the sequence differences in the subunit interface of avidin and AVR1–AVR7 would not interfere with tetramer formation. The most critical sequence difference is located at position 96, where methionine in avidin and AVR4/5 is replaced by lysine in AVR1–AVR3 and AVR6–AVR7. In avidin and AVR4/5, $Met^{96}$ interacts with $Met^{96}$ from a second monomer of the tetramer (Figure 2A). When $Met^{96}$ is replaced by the positively charged lysine residue, as seen in AVR1–AVR3 and AVR6–AVR7, it would be logical to expect that charge repulsion interferes with the formation of the tetramer. However, this does not seem to occur, but, instead, $Lys^{96}$ from monomer 1 can form hydrogen bonds with the side-chain hydroxy group of $Thr^{113}$ from monomer 4 and with the main-chain oxygen atom of $Val^{115}$ from monomer 3 (Figure 2B).

As compared with avidin, the biotin-binding pocket is highly conserved in AVR1–AVR7, where only minor changes in the sequences – and therefore in the shape of the binding pocket – are observed. These changes are limited to the area surrounding the most flexible part of the biotin molecule, and thus biotin can easily compensate for these changes. The only exception occurs at position 111. In avidin, AVR1 and AVR3–AVR7, lysine is found at this position, whereas, in AVR2, isoleucine is present. In the avidin crystal structure, the hydrophobic part of the lysine side chain interacts with the indole ring of $Trp^{110}$ keeping its position fixed (Figure 2C). The replacement of $Lys^{111}$ by the bulky isoleucine side chain would force the side chain of $Trp^{110}$ either to bend towards the binding pocket (effectively blocking biotin entry to the site) or away from the binding pocket (allowing the entry and exit of biotin into/from the site) (Figure 2D). $Trp^{110}$ is one of the most critical amino acids in the binding pocket, because it conforms exactly to the shape of biotin where the ligand is most rigid. The N-linked glycosylation site on L5, although close to the biotin-binding site, is solvent-exposed. The asparagine that is glycosylated is not in the vicinity of the ligand, whereas the serine side chain is hydrogen-bonded to the carboxylate group of biotin. When biotin is absent (for example, during post-translational glycosylation) the serine side chain should be accessible to the glycosylating enzyme.

All AVRs, with the exception of AVR2, contain a third single cysteine residue (in addition to the two present in avidin) located on the surface of the proteins. Since AVR2 does not have the extra cysteine residue, there is no reason to believe that AVR2 could form tetramers in a way different from that by which

**Table 2 Optical biosensor data for biotin and 2-iminobiotin binding for avidin and AVRs**

Shown below are the biotin-binding properties of AVRs compared with those of avidin (AVD), determined using the IAsys biosensor. The values for avidin are from Laitinen et al. [19] and Marttila et al. [17]. Reversibility of binding was determined by using a biotin cuvette. ND, not determined.

| | Value | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | AVD | AVR1 | AVR2 | AVR3 | AVR4/5 | AVR6 | AVR7 |
| $K_d$ (M)‡ | $(1.7 \pm 1.3) \times 10^{-8}$* | $(1.0 \pm 0.8) \times 10^{-7}$† | $(1.7 \pm 1.5) \times 10^{-6}$† | $\ll 10^{-8}$† | $(5.7 \pm 2.5) \times 10^{-7}$* | $\ll 10^{-8}$† | $(4.5 \pm 6.2)\ 10^{-9}$† |
| $K_d$ (M)§ | $(2 \pm 0.9) \times 10^{-8}$* | $(4.4 \pm 1.9) \times 10^{-8}$† | $(5.2 \pm 1.7) \times 10^{-8}$† | ND | $(9.1 \pm 3.6) \times 10^{-8}$* | ND | ND |
| $K_{ass}$ ($M^{-1} \cdot s^{-1}$) | $(1.6 \pm 0.7) \times 10^{4}$ | $(5.5 \pm 1.5) \times 10^{4}$ | $(1.8 \pm 0.1) \times 10^{4}$ | ND | $(1.8 \pm 0.2) \times 10^{5}$ | ND | $(1.8 \pm 0.2) \times 10^{5}$ |
| $K_{diss}$ ($s^{-1}$) | $(3.1 \pm 1.4) \times 10^{-4}$ | $(2.4 \pm 0.8) \times 10^{-3}$ | $(4.6 \pm 1.1) \times 10^{-3}$ | ND | $(1.7 \pm 0.7) \times 10^{-3}$ | ND | ND |
| Reversibility (%) | $1 \pm 1$ | $18 \pm 10$ | $94 \pm 3$ | $3 \pm 2$ | $2 \pm 2$ | $5 \pm 3$ | $3 \pm 3$ |

* Measured in a 2-iminobiotin cuvette.
† Measured in a biotin cuvette.
‡ Dissociation constants were calculated from the equilibrium-response data.
§ The dissociation constants were calculated directly from the binding curves.

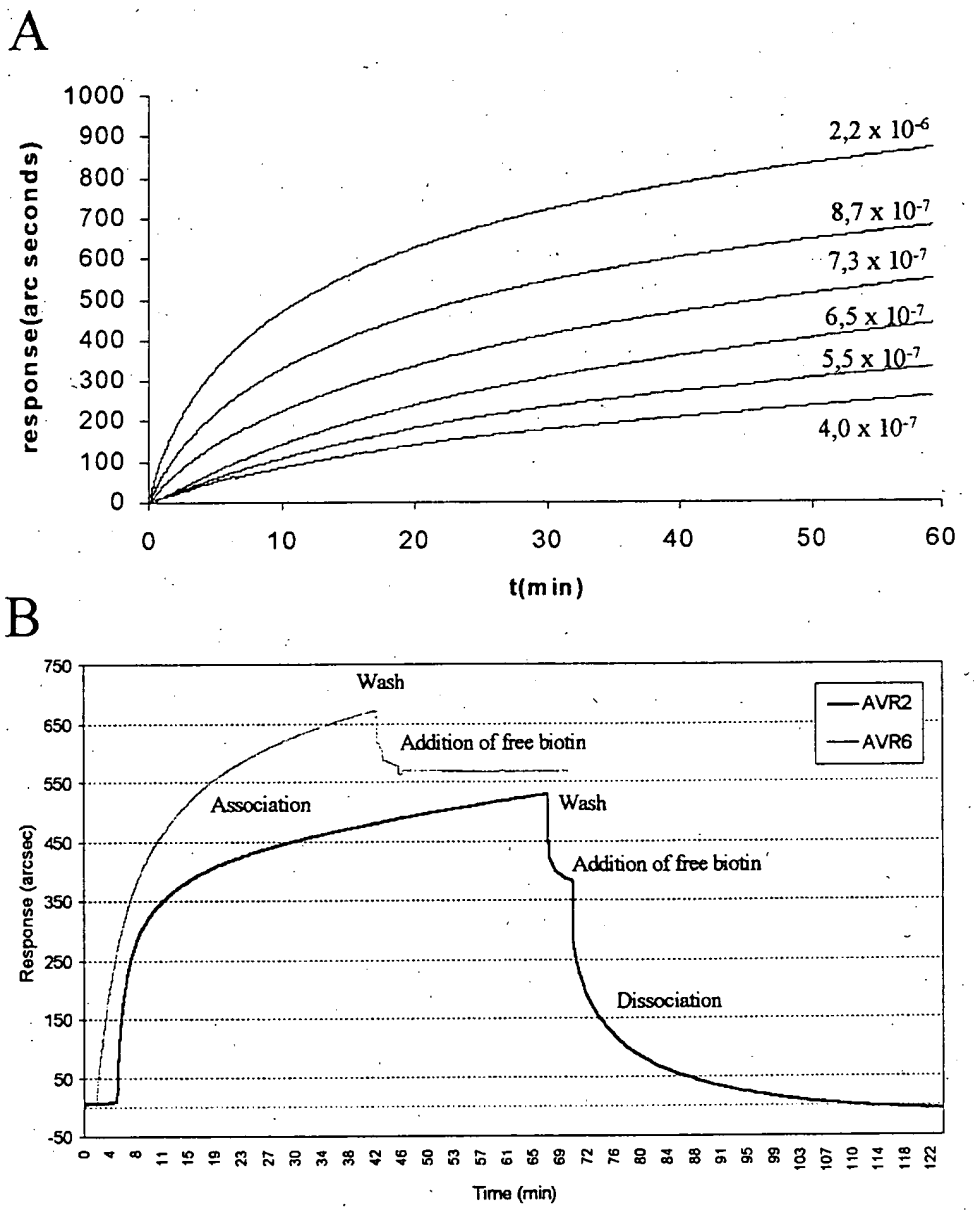


**Figure 5 (A) Binding curves for 2-iminobiotin binding by AVR4/5 at different concentrations, and (B) an example of an IAsys reversibility experiment**

(A) The first 500 s were used to calculate $k_{ass}$. The equilibrium state was obtained after measurement for about 1 h, depending on the concentration. The maximal binding ($R_{max}$) for the highest concentration was approx. 1000 arc-s. Note: 2,2 = 2.2 (etc.). (B) The protein was allowed to bind on to a biotin–aminosilane cuvette. After reaching equilibrium, the cuvette was washed and dissociation was measured. A surplus of free biotin was added and the measurement continued to achieve steady state. The cuvette was washed again. The values measured after binding and dissociation, and after biotin treatment and second washing step, were used to define the reversibility of the biotin binding. In this particular experiment, the reversibility was practically 0% for AVR6 and 100% for AVR2.

avidin does. The model structure for AVR4/5 (Figure 3) places the additional cysteine residue at position 124 (avidin numbering), close to the C-terminus. It is logical to expect two monomers to optimize their domain–domain interactions by forming a disulphide-bond between these unpaired cysteine residues (arrow in Figure 3), which, in the model structure of AVR4/5, are located near each other. The presence of an intermonomer disulphide bond at this location is not expected to interfere with tetramer formation. However, in AVR1, AVR3, AVR6 and AVR7, the extra cysteine residue is located in L4 at position 60 (avidin numbering), where it is impossible to form an extra disulphide bond between subunits without severely disturbing tetramer formation.

### Protein expression and purification

The AVR proteins were successfully produced in Sf9 insect cells. However, all AVRs showed different patterns on SDS/PAGE as compared with avidin [22] (Figure 4). One-step 2-iminobiotin (AVR1, AVRs 3–7) – and/or biotin (AVRs 1 and 2) – agarose affinity chromatography yielded highly homogenous proteins showing no contaminants as judged by SDS/PAGE analysis.

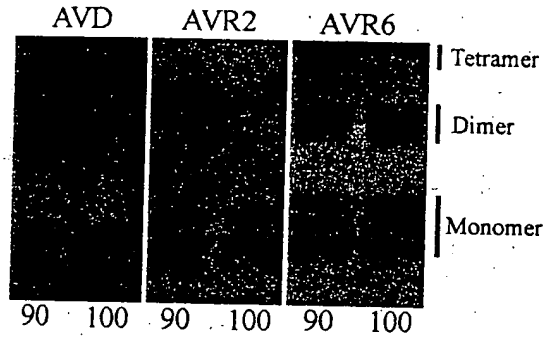


**Figure 6 Non-reducing SDS/PAGE of avidin (AVD), AVR2 and AVR6 at 90 and 100 °C**

The tetrameric, dimeric and monomeric forms are indicated with bars on the right side of the Figure.

### Biotin-binding properties

In the reversibility assay, AVRs 3–7 showed totally irreversible biotin binding, similarly to avidin. In contrast, AVR1 exhibited 18 and AVR2 93 % reversibility following addition of free biotin (Table 2). The actual dissociation constants ($K_d$) were calculated for the AVRs according to the $k_{ass}$ (association) and $k_{diss}$ (dissociation) rate constants measured in a separate assay at different protein concentrations (varying between 10 $\mu$M and 5 nM) (Table 2 and Figure 5). The binding seemed to follow second-order or even more complex kinetics. However, the first 1000 s of the binding fitted rather well to the biphasic curve. All measured $k_{diss}$ values followed the first-order kinetics. Because of the extremely high binding, affinities of the AVRs for biotin could not be determined except for AVR1, AVR2 and AVR7. Binding to 2-iminobiotin was determined only for AVR4/5, and it was similar to that of avidin. The lack of binding for the other AVRs was surprising, since most of them were purified using 2-iminobiotin–agarose. This may be due to the relatively short linker between 2-iminobiotin and the activated group of the IAsys cuvette that did not allow the 2-iminobiotin to penetrate deep enough into the biotin-binding pocket of the other AVRs.

### Structural analyses

All AVRs exhibited heat-stability comparable with avidin in a reducing SDS/PAGE assay (results not shown). In the absence of biotin, AVR tetramers began to dissociate into monomers at around 60 °C. In the presence of biotin, a portion of AVRs remained tetrameric even upon boiling. Non-reducing SDS/PAGE showed that AVR1 and AVRs 3–7 have a tendency to form dimers. Under reducing conditions, only tetramers and monomers could be discerned. This result suggests that the dimers are cross-linked via disulphide bridges (stable enough to hold the dimers together under non-reducing conditions) and are slowly degraded into monomers by heat (Figure 6 and Table 3). The shift from a fully tetrameric state into a dimeric and monomeric state was detected upon raising the temperature gradually. AVR2 is an exception, since it disintegrated directly into monomers, similarly to native avidin.

The AVRs also showed remarkable resistance against proteolysis. In the presence of biotin, all AVRs remained intact after 16 h of proteinase K treatment. Proteolysis occurred, albeit slowly, in the absence of biotin. Taken together, AVRs showed stability similar to, or even greater than, that of avidin (results not shown).

**Table 3 Comparative thermostability of avidin and AVRs under non-reducing conditions**

Shown below are the relative proportions of different oligomeric forms of AVRs and avidin (AVD) in SDS/PAGE under non-reducing conditions, boiled 20 min before loading on to the gel. Note the presence of dimeric forms in the case of AVR1 and AVR3–7.

| | Relative proportion of oligomeric forms (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| Form | AVD | AVR1 | AVR2 | AVR3 | AVR4/5 | AVR6 | AVR7 |
| Tetrameric | – | 20 | – | – | – | – | – |
| Dimeric | – | 40 | – | 50 | 50 | 50 | 20 |
| Monomeric | 100 | 40 | 100 | 50 | 50 | 50 | 80 |

**Table 4 Monoclonal anti-avidin antibody and polyclonal rabbit anti-avidin antibody ELISAs for avidin (AVD) and AVRs**

The values are means of two independent measurements of $A_{405}$ nm, after 60 min colour reaction.

| | $A_{405}$ | | | | | | |
|---|---|---|---|---|---|---|---|
| Antibody | AVR1 | AVR2 | AVR3 | AVR4/5 | AVR6 | AVR7 | AVD |
| Monoclonal anti-avidin | | | | | | | |
| Clone 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.66 |
| Clone 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0.64 |
| Polyclonal anti-avidin | 0.22 | 0.02 | 0.16 | 0.40 | 0.38 | 0.35 | 0.82 |

The multiple bands observed for the AVRs in SDS/PAGE were found to be differently glycosylated forms. Treatment with Endo $H_f$ (Figure 5) and PNGase F (results not shown) eliminated the higher-molecular-mass bands. The number of putative glycosylation sites seems to correlate well with the observed glycosylation patterns. Preliminary deglycosylation results using non-denaturing conditions suggest that N-glycan is attached to the glycosylation site in L5 in AVRs 1,2,4/5 (not shown). Isoelectric focusing showed that the pI values for the AVRs were approximately the same as expected from theoretical calculations, namely $\approx$ 7 for AVRs 1, 6 and 7, $\approx$ 5 for AVR2, and $\approx$ 10 for AVRs 3, and 4/5.

### Immunological analyses

Polyclonal rabbit anti-avidin antibody recognized AVRs 4/5–7 more weakly than it did avidin. Furthermore, the recognition of AVRs 1 and 3 was even more diminished, and AVR2 was not recognized at all. The two monoclonal anti-avidin antibodies tested did not recognize any of the AVRs (Table 4).

## DISCUSSION

The motivation for the present study was basically bipartite. First, we wanted to produce recombinant AVR proteins to examine their biochemical and functional properties, such as biotin binding, stability and immunological properties. Advantageous properties might be utilized to improve the current (strept)avidin–biotin technologies. Secondly, we wanted to reveal in detail the structural basis for the differences between avidin and the AVRs, which was addressed by careful sequence and modelling analyses.

Despite the relatively high primary sequence conservation [25–27] (Figure 1), there are interesting amino acid substitutions that make the physico-chemical properties of AVRs different from those of avidin. For example, the present study showed that the pI of some AVRs is neutral or even acidic, in contrast with the basic pI of avidin, and their glycosylation patterns also differ. Furthermore, differences in biotin binding were observed, regardless of the fact that almost all of the amino acid residues important for biotin binding in avidin [5] and streptavidin [3] are conserved in the AVRs. In previous studies, we have used the AVR sequences as models for lowering the pI of avidin [16] and to produce non-glycosylated avidin [17]. All of the pI-mutants, as well as the non-glycosylated avidins, bind biotin with high affinity and form stable tetramers. Therefore differences in pI or the lack of glycosylation at positions corresponding to residue 17 in avidin were not expected to affect biotin binding and tetramer formation of the AVRs. Consequently, the observed differences must be based on as-yet-unknown features that, while affecting biotin binding, do not impede tetramer stability.

Three biotin-bonding residues in loop L3 are altered in AVRs when compared with avidin. Two of these hydrogen-bonding interactions could be maintained despite having differing side chains at these positions in the AVRs. However, the hydrogen bond to biotin from the side chain of $Thr^{38}$ in avidin [5] is most likely lost when the threonine residue is replaced by alanine. By comparison, streptavidin shows only one hydrogen bond to biotin in loop L3 [3,4]. Therefore the differences in L3, including the presence of additional carbohydrate on some of the AVRs, may not be detrimental for biotin binding. The greatest functional differences were observed for AVR2, where binding to 2-iminobiotin was totally abolished and binding to biotin was virtually reversible. This result is in good agreement with the modelling results that suggest that the conversion of $Lys^{111}$ into isoleucine would alter the shape of the binding pocket.

The most interesting glycosylation site found in AVRs is located in L5. Residues of this loop, including the hydrophobic residues $Trp^{70}$ and $Phe^{72}$, as well as two hydrogen-bond-forming serine residues, contact biotin. However, AVR4/5 also has a glycosylation site located in L5, and still exhibits 2-iminobiotin binding as high as that shown by avidin. Therefore it seems that, in itself, glycosylation at this site does not affect biotin binding. The contribution of glycosylation to the reduced biotin binding seen in the other AVRs cannot be ruled out. Additional carbohydrate may also explain why none of the AVRs were recognized by the monoclonal anti-avidins antibodies: the epitope(s) could be masked by the sugar moieties in the AVRs.

Another structurally interesting feature of the AVRs (except for AVR2) is the third cysteine residue. The non-reducing-SDS/PAGE results, as well as molecular modelling, suggest that this extra cysteine residue forms an intermonomeric disulphide bridge, leading to SDS/PAGE patterns consistent with dimer formation. Cross-linking of monomers is most likely to occur in AVR4/5. In contrast, other cysteine-bearing AVRs either form bigger aggregates or their quaternary structures are different from that of avidin. Their inability to bind 2-iminobiotin in IAsys analyses may reflect the presence of such structural differences.

A $His^{127}$ → aspartate mutation introduced into the subunit interface of streptavidin prevented tetramer formation through charge repulsion [28]. Mutation of all of the residues at the 1–3 (and 2–4) interface of avidin into alanine residues also showed that the stability of the resultant mutant was reduced [19]. However, the $Met^{96}$ → lysine substitution does not appear to interfere with tetramer formation in the AVRs, but may function to stabilize the tetramer (Figure 2). At the large 1–4 (and 2–3) interface, the substitution of $Asn^{54}$ and $Asn^{69}$ by histidine residues is especially interesting [5,13]. The two asparagine residues participate in a network of ten hydrogen bonds in avidin [5] at both ends of the 1–4 (2–3) dimer. Thus these two residues are involved in networks of 20 hydrogen bonds between subunits 1 and 4 (and 2 and 3) [18]. However, histidine is also capable of forming hydrogen bonds and may functionally substitute for asparagine.

There are several possible explanations for the stability of the AVRs. Substitutions at the 1–4 (2–3) interface can be complementary; while an amino acid change in one subunit decreases the interface affinity, the coincident mutation in the other subunit may restore it. Bogan and Thorn [29] have introduced a 'hot spot' model for interface contacts, where they propose that the free energy of interface binding is not evenly distributed, but is concentrated at hot spots consisting of small subsets of residues. A hot spot contains energetically important key amino acids (commonly hydrophobic) surrounded by energetically unimportant residues (called 'O-ring' residues), whose effect is to exclude bulk solvent from the hot spot. Thus the energetically less important residues may be mutated frequently without causing major effects on interface affinity and stability. Furthermore, water molecules often play an essential role in protein–protein interactions [30]. The conversion of bulky charged or polar residues into smaller ones may result in a situation where a water molecule functions to bridge the hydrogen bond previously formed directly between the bulky residue and its counterpart. Therefore the subunit interfaces of the AVRs can tolerate mutations within the O-ring zone, and some substitutions may allow water molecules in the interface without seriously affecting interface affinity.

Avidin is generally thought to act as an antimicrobial agent by depriving invading micro-organisms of biotin. It is therefore possible that the different AVRs exist to broaden the range of host defence. Recently, Zerega and co-workers [31] found that avidin is expressed in skeletal muscle and growth plate hypertrophic cartilage of the developing chicken embryo. This finding, together with their *in vitro* results suggest that by interfering with fatty acid metabolism avidin assists the terminal differentiation of chondrocytes and myoblasts. Similarly, the fibropellins expressed during ontogenesis of sea urchins contain a domain similar to that of avidin [32,33]. Fibropellins may form dimers or higher-order oligomeric structures and thereby promote protein–protein interactions during embryogenesis. The AVRs in chicken may offer comparable oligomeric scaffolds for embryonic development and/or host-defence functions.

In conclusion, the recombinant AVRs are functional and show interesting physico-chemical properties. These new biotin binders may provide advantages over avidin and streptavidin in several applications [34]. For example, AVR2 could be used in affinity-purification protocols that require mild elution conditions. Nonetheless, further analyses are required to solve the quaternary structures, precise glycosylation compositions and immunological properties of the AVRs in detail.

We thank Mrs Irene Helkala, Mrs Pirjo Käpylä, Mr Paavo Niutanen and Mr Jarno Hörhä for first-class technical assistance. This study has been supported by grants from the Academy of Finland.

## REFERENCES

1 Wilchek, M. and Bayer, E. A. (1999) Foreword and introduction to the book (Strept)avidin–Biotin System. Biomol. Eng. **16**, 1–4

2 Bayer, E. A. and Wilchek, M. (1990) Application of avidin–biotin technology to affinity-based separations. J. Chromatogr. **510**, 3–11

3 Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J. and Salemme, F. R. (1989) Structural origins of high-affinity biotin binding to streptavidin. Science **243**, 85–88
4 Hendrikson, W. A., Pähler, A., Smith, J. L., Satow, Y., Merritt, E. A. and Phizackerley, R. P. (1989) Crystal structure of core streptavidin determined from multiwavelength anomolous diffraction of synchrotron radiation. Proc. Natl. Acad. Sci. U.S.A. **86**, 2190–2194
5 Livnah, O., Bayer, E. A., Wilcheck, M. and Sussman, J. L. (1993) Three-dimensional structures of avidin and the avidin–biotin complex. Proc. Natl. Acad. Sci. U.S.A. **90**, 5076–5080
6 Pugliese, L., Coda, A., Malcovati, M. and Bolognesi, M. (1993) Three-dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at 2.7 Å resolution. J. Mol. Biol. **231**, 698–710
7 Green, N. M. (1990) Avidin and streptavidin. Methods Enzymol. **184**, 51–67
8 Wang, C., Eufemi, M., Turano, C. and Giartosio, A. (1996) Influence of the carbohydrate moiety on the stability of glycoproteins. Biochemistry **35**, 7299–7307
9 Gonzalez, M., Argarana, C. E. and Fidelio, G. D. (1999) Extremely high thermal stability of streptavidin and avidin upon biotin binding. Biomol. Eng. **16**, 67–72
10 Green, N. M. (1975) Avidin. Adv. Prot. Chem. **29**, 85–133
11 Keinänen, R. A., Laukkanen, M. L. and Kulomaa, M. S. (1988) Molecular cloning of three structurally related genes for chicken avidin. J. Steroid Biochem. **30**, 17–21
12 Keinänen, R. A., Wallen, M. J., Kristo, P. A., Laukkanen, M. O., Toimela, T. A., Helenius, M. A. and Kulomaa, M. S. (1994) Molecular cloning and nucleotide sequence of chicken avidin-related genes 1–5. Eur. J. Biochem. **220**, 615–621
13 Ahlroth, M. K., Kola, E. H., Ewald, D., Masabanda, J., Sazanov, A., Fries, R. and Kulomaa, M. S. (2000) Characterization and chromosomal localization of the chicken avidin gene family. Anim. Genet. **31**, 367–375
14 Ahlroth, M., Alhroth, P. and Kulomaa, M. S. (2001) Copy-number fluctuation by unequal crossing-over in the chicken avidin gene family. Biochim. Biophys. Res. Commun. **26**, 400–406
15 Kunnas, T. A., Wallen, M. J. and Kulomaa, M. S. (1993) Induction of chicken avidin and related mRNAs after bacterial infection. Biochim. Biophys. Acta **1216**, 441–445
16 Marttila, A. T., Airenne, K. J., Laitinen, O. H., Kulik, T., Bayer, E. A., Wilchek, M. and Kulomaa, M. S. (1998) Engineering of chicken avidin: a progressive series of reduced charge mutants. FEBS Lett. **441**, 313–317
17 Marttila, A. T., Laitinen, O. H., Airenne, K. J., Kulik, T., Bayer, E. A., Wilchek, M. and Kulomaa, M. S. (2000) Recombinant NeutraLite avidin: a non-glycosylated, acidic mutant of chicken avidin that exhibits high affinity for biotin and low non-specific binding properties. FEBS Lett. **467**, 31–36
18 Laitinen, O. H., Airenne, K. J., Marttila, A. T., Kulik, T., Porkka, E., Bayer, E. A., Wilchek, M. and Kulomaa, M. S. (1999) Mutation of a critical tryptophan to lysine in avidin or streptavidin may explain why sea urchin fibropellin adopts an avidin-like domain. FEBS Lett. **461**, 52–58
19 Laitinen, O. H., Marttila, A. T., Airenne, K. J., Kulik, T., Livnah, O., Bayer, E. A., Wilchek, M. and Kulomaa, M. S. (2001) Biotin induces tetramerization of a recombinant monomeric avidin. A model for protein–protein interactions. J. Biol. Chem. **276**, 8219–8224

Received 5 October 2001/17 January 2002; accepted 22 February 2002

20 Johnson, M. S. and Overington, J. P. (1993) A structural basis for sequence comparisons. An evaluation of scoring methodologies. J. Mol. Biol. **233**, 716–738
21 Johnson, M. S., May, A. C., Rodionov, M. A. and Overington, J. P. (1996) Discrimination of common protein folds: application of protein structure to sequence/structure comparisons. Methods Enzymol. **266**, 575–598
22 Airenne, K. J., Oker-Blom, C., Marjomäki, V. S., Bayer, E. A., Wilchek, M. and Kulomaa, M. S. (1997) Production of biologically active recombinant avidin in baculovirus-infected insect cells. Protein Expression Purif. **9**, 100–108
23 Bayer, E. A., Ehrlich-Rogozinski, S. and Wilchek, M. (1996) Sodium dodecyl sulfate–polyacrylamide gel electrophoretic method for assessing the quaternary state and comparative thermostability of avidin and streptavidin. Electrophoresis **17**, 1319–1324
24 Engvall, E. and Perlmann, P. (1972) Enzyme-linked immunosorbent assay, ELISA. III. Quantitation of specific antibodies by enzyme-labelled anti-immunoglobulin in antigen coated tubes. J. Immunol. **109**, 129–135
25 Delange, R. J. and Huang, T. S. (1971) Egg white avidin. III. Sequence of the 78-residue middle cyanogen bromide peptide: complete amino acid sequence of the protein subunit. J. Biol. Chem. **246**, 698–709
26 Gope, M. L., Keinänen, R. A., Kristo, P. A., Conneely, O. M., Beattie, W. G., Zarucki-Schulz, T., O'Malley, B. W. and Kulomaa, M. S. (1987) Molecular cloning of the chicken avidin cDNA. Nucleic Acids Res. **15**, 3595–3606
27 Ahlroth, M. K., Grapputo, A., Laitinen, O. H. and Kulomaa, M. S. (2001) Sequence features and evolutionary mechanisms in the chicken avidin gene family. Biochem. Biophys. Res. Commun. **285**, 734–741
28 Sano, T., Vajda, S., Smith, C. L. and Cantor, C. R. (1997) Engineering subunit association of multisubunit proteins: A dimeric streptavidin. Proc. Natl. Acad. Sci. U.S.A. **94**, 6153–6158
29 Bogan, A. A. and Thorn, K. S. (1998) Anatomy of hot spots in protein interfaces. J. Mol. Biol. **280**, 1–9
30 Janin, J. (1999) Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition. Structure Fold. Des. **7**, R277–R279
31 Zerega, B., Camardella, L., Cermelli, S., Sala, R., Cancedda, R. and Descalzi Cancedda, F. (2001) Avidin expression during chick chondrocyte and myoblast development *in vitro* and *in vivo*: regulation of cell proliferation. J. Cell Sci. **114**, 1473–1482
32 Bisgrove, B. W. and Raff, R. A. (1993) The SpEGF III gene encodes a member of the fibropellins: EGF repeat-containing proteins that form the apical lamina of the sea urchin embryo. Dev. Biol. **157**, 526–538
33 Bisgrove, B. W., Andrews, M. E. and Raff, R. A. (1995) Evolution of the fibropellin gene family and patterns of fibropellin gene expression in sea urchin phylogeny. J. Mol. Evol. **41**, 34–45
34 Chinol, M., Casalini, P., Maggiolo, M., Canevari, S., Omodeo, E. S., Caliceti, P., Veronese, F. M., Cremonesi, M., Chiolerio, F., Nardone, E. et al. (1998) Biochemical modifications of avidin improve pharmacokinetics and biodistribution, and reduce immunogenicity. Br. J. Cancer **78**, 189–197

Egg albumin (Sigma) and *Pertussis* (Connaught Laboratories) were administered according to protocol described in M. H. Perdue, M. Chung, D. G. Gall, *Gastroenterology* 86, 391 (1984).

11. S. J. King *et al.*, *Eur. J. Immunol.* 16, 151 (1986); S. J. King and H. R. P. Miller *Immunology* 51, 653 (1984).
12. K. J. Bloch *et al.*, *Gastroenterology* 77, 1039 (1979); E. Jarrett and H. Bazin, *Nature* 251, 613 (1974); *Clin. Exp. Immunol.* 30, 330 (1977).
13. Animals were removed from the colony room and placed in plastic cages inside concrete-encased (soundproof) cabinets. A light flashed at an alteration rate of 300 ms, and background noise was provided by ventilation fans. The AV CS was based on that used by G. MacQueen and S. Siegel (*Behav. Neurosci.*, in press).
14. An enzyme-linked immunosorbent assay (ELISA) for detecting RMCP II was modified from Miller *et al.* (7). Rats were anesthetized with ether, and blood was obtained from the retro-orbital plexus. Sera were collected and stored at −20°C. The wells of a tissue culture microtitre plate (Nunclon Delta) were coated with 0.5 mg of RMCP II per milliliter of a 0.2*M* carbonate buffer, *p*H 9.6. Samples and standards were diluted in PBS containing 0.3% w/v bovine serum albumin, 0.02% v/v polyoxyethylene-sorbitan monolaurate (Tween 20), and 0.02% w/v sodium azide and incubated for 16 to 24 hours with a diluted specific rabbit antiserum to RMCP II (anti-RMCP II) from which all activity against RMCP I had been removed by immunoabsorption. After extensive washing of the plates, 100 μl of samples and standards were placed in duplicate wells and incubated for 16 to 24 hours. Plates were again washed several times and rabbit antibody bound to the plate was detected with an alkaline phosphatase conjugated goat to rabbit antibody (ICN) and a sodium-*p*-nitrophenyl phosphate substrate (Sigma). Results were calculated on the basis of a standard curve constructed with the use of known concentrations of purified RMCP II. Antibodies to RMCP II were raised in rabbits with purified RMCP II. Antisera were absorbed twice with RMCP I covalently attached to Sepharose 4B. The final preparation showed strong binding to RMCP II on immunoblot analysis, with a weak cross-reaction to RMCP I, and no detectable binding to cathepsin G. The specificity of this assay was confirmed with homogenates of tongue tissue known to contain high levels of RMCP I, but no RMCP II. Although there was some binding to RMCP I on the immunoblot, no binding was detected in the liquid phase.
15. Morphological studies have shown the presence of mast cells in peripheral nerves [Y. Olsson, *Acta. Neurol. Scand.* 47, 357 (1971)] and autonomic ganglia [G. Gabella, *Structure of the Autonomic Nervous System* (Chapman and Hall, London, 1976)], a consistent ultrastructural relationship between mucosal mast cells and nerves in normal and nematode-infected rat lamina propria was detected [R. H. Stead, *Proc. Natl. Acad. Sci. U.S.A.* 84, 2975 (1987)].
16. Substance P causes release of histamine from mast cells in vitro [F. Shanahan *et al.*, *J. Immunol.* 135, 1331 (1985); O. Hagermark, T. Hockfelt, B. Pernow, *J. Invest. Dermatol.* 71, 233 (1978)], and mast cells and substance P containing nerves may be involved in the vasodilatory response to noxious stimuli [J. C. Foreman and C. C. Jordan, *J. Physiol.* 238, 58 (1982)].
17. Evidence has also supported a role for functional nerve–mast cell interactions. A. R. Leff *et al.* [*J. Physiol.* 136, 1066 (1986)] demonstrated that vagal stimulation causes enhanced histamine release from mast cells after antigen challenge, and a decrease in mast cell granularity has been shown after electrical field stimulation [T. Bani-Sacchi *et al.*, *J. Physiol.* 371, 29 (1986)]. Studies of hypersensitivity reactions in the gut and lung indicate a neural component in the changes in epithelial ion transport induced by antigen [Y. Harari *et al.*, *J. Immunol.* 138, 1250 (1987); see M. H. Perdue in (*10*)].
18. D. Befus, F. Pearce, J. Bienenstock, in *Food Allergy and Intolerance*, J. Brostoff and S. J. Challacombe, Eds. (Bailliere-Tindall, London, 1986), pp. 88–98; E. G. Seidman, D. G. Hanson, W. A. Walker, *Gastroenterology* 90, 120 (1986); J. B. Kirsner and R. G. Shorter, *N. Engl. J. Med.* 306, 837 (1982).
19. J. C. Foreman, *Allergy* 42, 1 (1987); D. M. Barnes, *Science 232*, 160 (1986); P. J. Barnes, *Lancet* i 242 (1986).
20. Supported by grants from the Medical Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, and the Foundation for Ileitis and Colitis. We thank L. Neilson for technical assistance and R. Woodbury for the RCMP II and the anti-RCMP II.

18 July 1988; accepted 17 October 1988

# Structural Origins of High-Affinity Biotin Binding to Streptavidin

PATRICIA C. WEBER, D. H. OHLENDORF, J. J. WENDOLOSKI, F. R. SALEMME*

**The high affinity of the noncovalent interaction between biotin and streptavidin forms the basis for many diagnostic assays that require the formation of an irreversible and specific linkage between biological macromolecules. Comparison of the refined crystal structures of apo and a streptavidin:biotin complex shows that the high affinity results from several factors. These factors include the formation of multiple hydrogen bonds and van der Waals interactions between biotin and the protein, together with the ordering of surface polypeptide loops that bury the biotin in the protein interior. Structural alterations at the biotin binding site produce quaternary changes in the streptavidin tetramer. These changes apparently propagate through cooperative deformations in the twisted β sheets that link tetramer subunits.**

STREPTAVIDIN IS A TETRAMERIC PROTEIN (molecular weight = 4 × 15,000) isolated from the actinobacterium *Streptomyces avidinii* (*1*). Streptavidin, and the homologous protein avidin, are remarkable for their ability to bind up to four molecules of *d*-biotin with unusually high affinity [dissociation constant $K_d = 10^{-15} M$ (*1, 2*)]. Although these proteins may function as antibiotics that deplete the environment of the essential vitamin biotin, they have been studied primarily as paradigms for understanding high-affinity protein-ligand interactions (*2*). At the same time, the ability of streptavidin and avidin to bind derivatized forms of biotin has led to their widespread use in diagnostic assays that require formation of an essentially irreversible and specific linkage between biological macromolecules (*3*). We undertook the structure determination of streptavidin, with and without bound biotin, to uncover the origins of high affinity of the protein for biotin.

Streptavidin was obtained from several commercial sources and produced different crystal forms during the course of the study. The most consistent results were obtained with a fragment of the native 159-residue streptavidin chain, incorporating residues 13 through 133. Numerous studies indicate that this truncated form of the molecule binds biotin with an affinity that is the same or similar to alternative longer versions of the protein. Moreover, in some cases it appeared that preparations identified as full-length material crystallized isomorphously with the truncated fragment, suggesting that the molecular terminii may be relatively flexible or disordered. Crystallization conditions for apostreptavidin and its biotin complex were found by robotic grid search methods (*4*). Both formed crystals from a polyethylene glycol–LiCl mixture, although the streptavidin:biotin complex crystallizes at *p*H 7.8 [space group $I4_122$, $a = b = 99.4$ Å, $c = 125.8$ Å (*5*)], whereas apostreptavidin crystallizes at *p*H 2.4 [space group $I4_122$, $a = b = 58.3$ Å, $c = 172.5$ Å]. Unit cell parameters of truncated apostreptavidin are similar to those reported by Pahler *et al.* (*6*), although the crystals studied here grow at lower *p*H, and diffract to higher resolution ($d_{min} = 1.7$ Å).

The structure of apostreptavidin was determined by multiple isomorphous replacement techniques. X-ray diffraction data for parent crystals and several isomorphous replacement derivatives were collected using a multiwire area detector and processed with the Xengen data-reduction package (*7*). Successful derivatives included $K_2Pt(SCN)_6$, which was prepared by soaking crystals in the heavy metal solution, and an iodine derivative prepared by crystallizing protein after reaction in solution (*8*). Substitution sites were located by an automated search procedure (*9*) performed on the heavy atom difference Patterson maps. Phases were ob-

Central Research & Development Department, E. I. du Pont de Neumours and Company, Inc., Du Pont Experimental Station, ES 228/320, Wilmington, DE 19880–0228.

*To whom correspondence should be addressed.

tained after refinement of heavy atom positions by the origin-removed different Patterson method (*10*) and used to compute a 2.6 Å resolution electron density map (*11*).

The electron density map was interpreted with the graphics program FRODO (*12*). Identification of several Trp and other large residues allowed an initial α-carbon backbone trace and sequence assignment to be made for ~70% of the structure, organized primarily as antiparallel β sheet. A partial atomic model was constructed from the α-carbon trace with the use of fragment superposition (*13*) from a database of refined protein structures (*14*). Complete backbone fragments that best fit the α-carbon trace were incorporated into the model. Side chains were positioned by displaying possible rotamers for each amino acid from a library (*15*) and by including coordinates for the rotamer that best fit the electron density. Phases computed from the partial model were combined with the multiple isomorphous replacement phases and used to generate an improved electron density map whose interpretation defined the remainder of the molecular structure. The structure was refined with the restrained least-squares method of Hendrickson and Konnert (*16*) and manual rebuildings in electron density maps. Although the refinement proceeded smoothly, two surface loops lacked defined density and appeared disordered in the final structure. The crystallographic *R*-factor for apostreptavidin, including residues 13 to 46, 49 to 63, and 69 to 133, as well as 33 water molecules, is 0.21 for 11,120 reflections [$F_{observed} > \sigma(F_{observed})$] between 5.0 and 1.8 Å resolution.

The structure of the streptavidin:biotin complex was solved with the use of symmetry-constrained searches of the complex unit cell with the apostreptavidin tetramer (*17*). Apostreptavidin crystallizes with a monomer in the asymmetric unit, so that subunits of the tetramer are related by crystallographic dyad axes. Since tetramers in the nonisomorphous crystals of the streptavidin:biotin complex could also pack with subunits related by crystal symmetry, we searched that cell using the apostreptavidin tetramer as the probe molecule (*18*). A maximum correlation coefficient of 0.56 was obtained for 1363 intensities between 4 and 5 Å resolution when the streptavidin tetramer was positioned at the origin of the streptavidin:biotin complex unit cell with all three of its dyad axes coincident with the crystallographic dyad axes. This result shows that both apo and liganded forms of streptavidin are tetramers with subunits related by 222 point group symmetry. The initial crystallographic *R*-factor for apostreptavidin positioned in the biotin:streptavidin unit cell was 0.41 for data from 5.0 to 2.6 Å resolution.

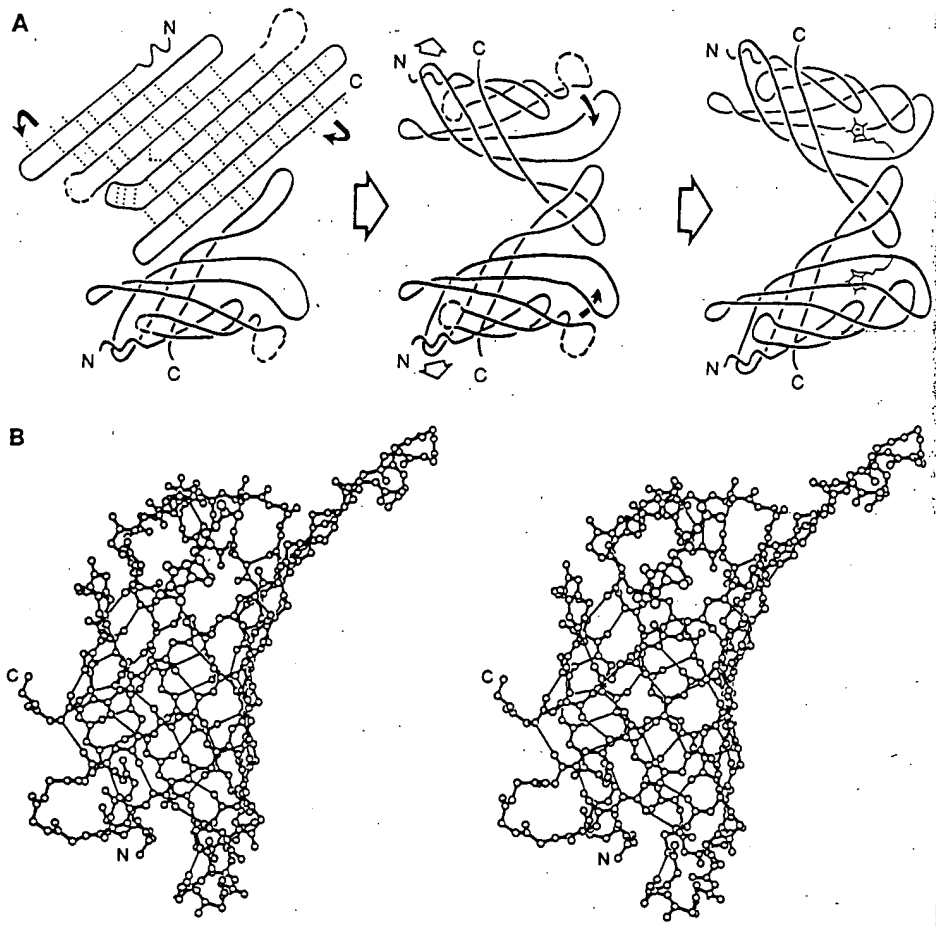


**Fig. 1.** Streptavidin structure. (**A**) Cartoon sequentially showing the β sheet folding pan of the hydrogen-bonded dimer, the apostreptavidin structure, and changes upon biotin binding. These changes include ordering of two loops (shown dashed) incorporating residues 45 to 50 and 63 to 69. (**B**) Stereoview of a streptavidin subunit with biotin bound, showing β barrel hydrogen bonds as thin lines. Residues 13 through 133 form an eight-stranded antiparallel β sheet wrapped as a slightly flattened barrel.

The structure of the streptavidin:biotin complex was refined by a combination of conventional restrained least-squares methods and crystallographically constrained molecular dynamics. The molecular dynamics refinement protocol essentially followed previous work (*19*), but was implemented in our laboratory by combining features of AMBER (*20*), PROLSQ (*16*), and PROFFT (*21*). The crystallographic *R*-factor of the streptavidin:biotin complex, including all residues between sequence positions 13 and 133, biotin, and nine water molecules, is 0.22 for 7379 reflections with $F_{observed} > \sigma(F_{observed})$ between 5.0 and 2.6 Å resolution [coordinates will be deposited in the Brookhaven Protein Data Bank (*14*)].

Streptavidin subunits are organized as eight-stranded, sequentially connected, antiparallel β sheets. The sheets are formed of coiled polypeptide chains with a staggered pattern of adjacent strand hydrogen-bond registration (*22*). This arrangement produces a cyclically hydrogen-bonded barrel with several extended hairpin loops, including one near the carboxyl terminus whose edge is free to form a more extended β sheet (Fig. 1). Pairs of streptavidin barrels hydrogen bond together at this free edge to form symmetric dimers that resemble basketball nets connected by their rims at a 45° angle. The naturally occurring streptavidin tetramer is formed by interdigitating a pair of dimers, with their dyad axes coincident, to produce a particle with 222 point group symmetry. The tetramer is stabilized by extensive van der Waals interactions between the subunit barrel surfaces, which have complementary curvatures (Fig. 2).

Biotin binds in pockets at the ends of each of the streptavidin β barrels (Fig. 1). The residues lining the pockets are primarily aromatic or polar amino acids or both. These groups are solvent exposed in apostreptavidin so that several water molecules occupy the biotin binding site. Biotin binding involves displacement of bound water, formation of multiple interactions between biotin heteroatoms and the binding site

residues, and burial of the biotin through ordering of a suface loop (residues 45 to 50) that is disordered in apostreptavidin. Polar interactions made between biotin heteroatoms and the protein include (i) an extensive pattern of hydrogen bonds with the biotin ureido group, where no less than five protein residues form associations; (ii) a possible interaction between biotin sulfur and the hydroxyl group of $Thr^{90}$; and (iii) hydrogen-bonded interactions with the valeryl carboxyl group that includes a hydrogen bond from the backbone NH of $Asn^{49}$, which becomes ordered on biotin binding (Fig. 3). Several other residues lining the binding site are immobilized by hydrogen bonds, which are formed in many cases with the same residues that hydrogen bond to biotin. These include Trp residues 79, 92, and 108 that pack around the biotin tetrahydrothiophene ring, and which, together with $Trp^{120}$ from the cyad-related subunit, form a hydrophobic biotin binding site. As a result of this extensive pattern of interactions, resulting in part from the ordering of loop 45 to 50, bound biotin is essentially buried in the complex with only the valeryl carboxyl oxygens partially accessible to solvent.

Experimental studies of the binding of biotin analogs to avidin, a tetrameric protein from avian egg white that shares 38% sequence identity (*23*) with the crystallographically defined streptavidin, suggest that interactions made with the ureido ring system predominate in stabilizing the biotin-protein complex (*2*). An unusual aspect of the interaction involves participation of the bio-

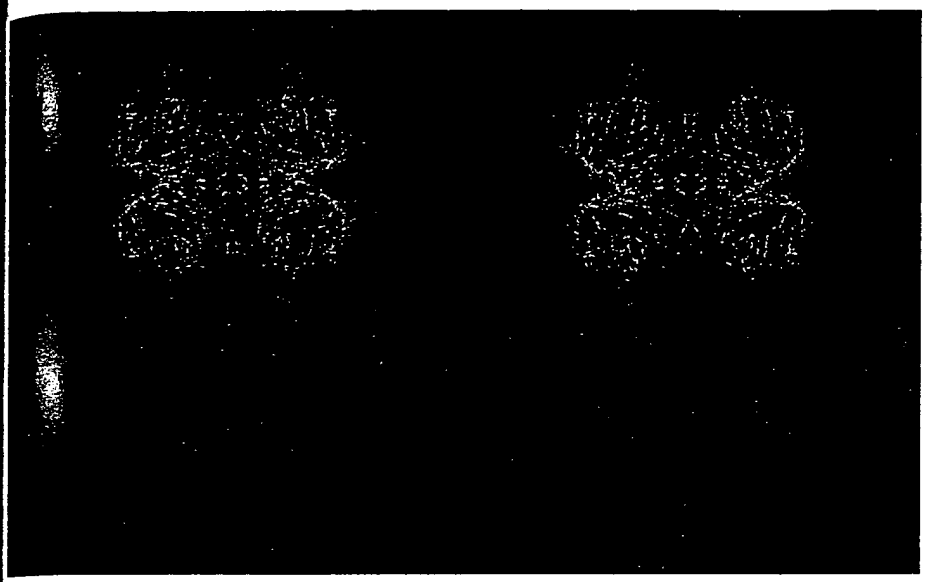

**Fig. 2.** Stereoviews of the streptavidin tetramer. (**Top**) Tetramer with bound biotins (backbone atoms), viewed along dyad symmetry axis relating hydrogen-bonded subunits (hydrogen bonds in red). (**Bottom**) Hydrogen-bond circuit representation of the tetramer, as defined by sheet hydrogen bonds and backbone amide group atoms. This representation emphasizes the continuity of the interactions that distribute forces throughout the barrels and across the subunit dimer axes.

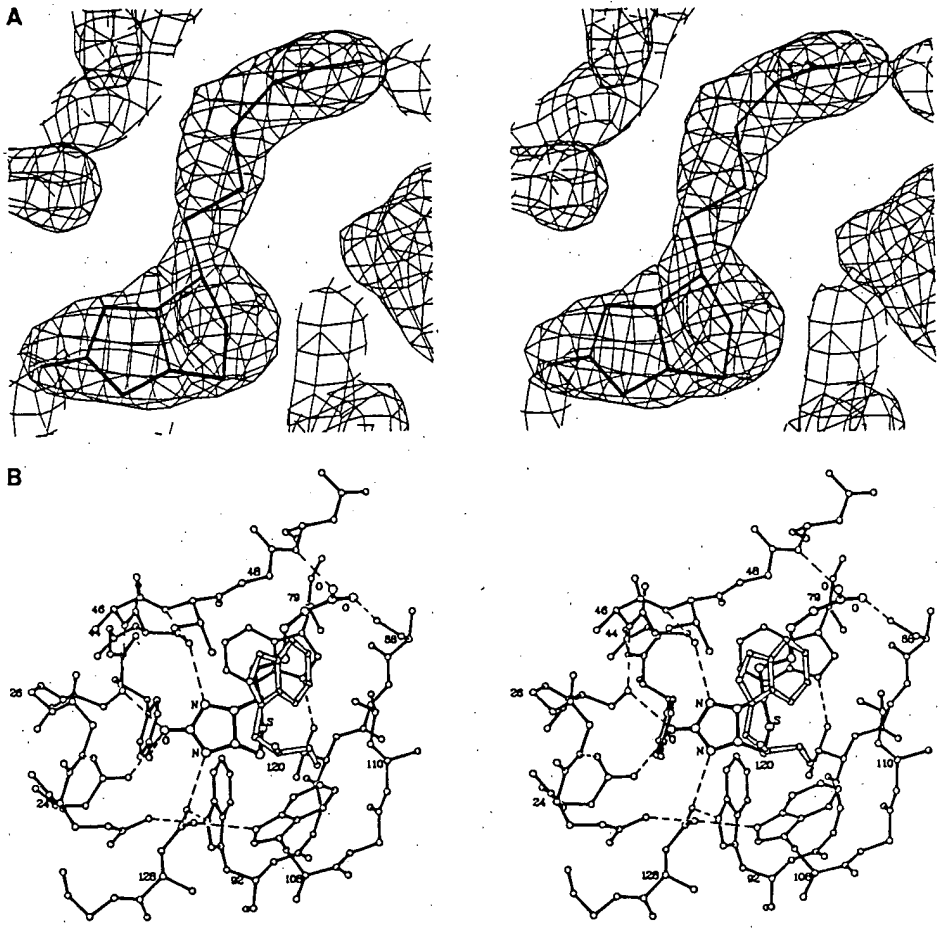


**Fig. 3.** Biotin binding site. (**A**) Biotin electron density from final $(2F_o - F_c)\alpha_{calc}$ 2.6 Å electron density map. (**B**) Biotin interactions. One indole ring (open bonds) is contributed by $Trp^{120}$ from the symmetry-related, hydrogen-bonded dimer. Residues shown include: $Asn^{23}$, Gln, Leu, Gly, $Ser^{27}$; $Tyr^{43}$, Glu, Ser, Ala, Val, Gly, $Asn^{49}$; $Trp^{79}$; $Ser^{88}$, Ala, Thr, Thr, $Trp^{92}$; $Trp^{108}$, Leu, $Leu^{110}$; $Asp^{128}$; and $Thr^{129}$. Some side chains are omitted for clarity. Dashed lines show the hydrogen-bonding pattern for residues that interact with biotin heteroatoms. The valeryl oxygens of biotin are hydrogen-bonded to the backbone NH of $Asn^{49}$ and Oγ1 of $Ser^{88}$, respectively. The $Thr^{90}$ side chain oxygen is near the biotin sulfur and $Trp^{79}$ Nε1. $Ser^{43}$ Oγ1 interacts with one biotin ureido NH (upper) and the backbone NH of $Val^{47}$. The other biotin ureido NH (lower) is hydrogen-bonded to a carboxyl oxygen of $Asp^{128}$. The $Asp^{128}$ side chain oxygens form additional interactions with $Trp^{92}$ Nε1, $Trp^{108}$ Nε1, and $Gln^{24}$ Nε1. Three side chain atoms, $Tyr^{43}$ OH, $Asn^{23}$ Nδ2H, and $Ser^{27}$ OH, are situated to hydrogen bond with the biotin ureido oxygen. The OH of $Ser^{27}$ is also hydrogen-bonded to $Ala^{46}$ NH, and $Asn^{23}$ Oδ1 interacts with $Leu^{25}$ NH. Note that residues forming hydrogen bonds with biotin ureido oxygen are themselves stabilized by orienting hydrogen bonds to backbone NH groups.

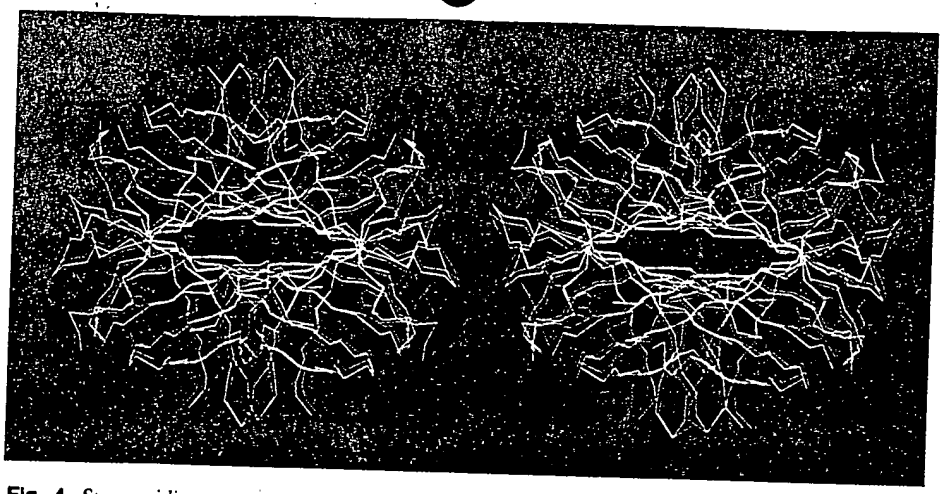

**Fig. 4.** Streptavidin quaternary structural changes. Yellow lines show α-carbon backbone trace of apostreptavidin and blue lines show streptavidin with biotin bound. Pairs of hydrogen-bonded dimers have been separated by 15 Å along the horizontal tetramer dyad axis for clarity. Changes in the hydrogen-bonded dimer geometry produce relative rotations of the dimers about the horizontal dyad. Tetramer subunits of both the apostreptavidin and streptavidin:biotin complex crystals are related by crystal-symmetry operations. The root-mean-square (rms) displacement for 112 common subunit Cα atoms, relative to the origin defined by the intersection of the tetramer 222 symmetry axes is 2.0 Å; the rms fit for the 67 Cα atoms in the β barrel is 0.7 Å in this frame. Superpositioning individual subunits gives an rms of 1.9 Å for all 112 Cα atoms common to the apo and liganded proteins and 0.3 Å for the 67 β barrel Cα atoms.

tin ureido group in an extended hydrogen-bond network anchored by the buried carboxyl group of $Asp^{128}$ that hydrogen bonds one ureido NH. The latter hydrogen bond could stabilize resonance forms that localize positive charge on biotin nitrogens and negative charge at the biotin ureido oxygen. Indeed, the ureido oxygen forms three hydrogen bonds, arranged with tetrahedral geometry, suggesting that the groups involved stabilize an $sp^3$ oxyanion (Fig. 3B). Comparison of streptavidin and avian avidin show that all of the groups that directly bind biotin are conserved with the exception of $Ser^{45}$, which is replaced by Thr with similar functionality, and $Asp^{128}$, which, surprisingly, is substituted by Asn (*23*). These changes may reflect some differences in the way biotin is stabilized in streptavidin and avidin. However, the analog of the residue that hydrogen bonds to $Asp^{128}$ in streptavidin, $Gln^{24}$, is substituted by Asp in avidin, so that slightly different but similar polarization networks could be functional in both proteins. Although biotin makes additional hydrophobic and hydrogen-binding interactions that assist binding, the hydrogen-bonded interactions with the valeryl group appear to play a lesser role. This group is partially accessible in the complex and provides the covalent attachment sites for linking biotin with other biomolecules (*3*).

Apo and liganded streptavidin differ in quaternary structure. Although the observed changes could in part reflect differences in crystal *p*H or lattice interactions and there is currently no evidence for subunit cooperativity, the pattern of quaternary changes nevertheless suggests a consistent mechanism of subunit communication. Subunit differences between apo and liganded streptavidin include the formation of extensive biotin:protein interactions, concomitant ordering of two surface loops, and formation of a salt link between $Glu^{51}$ and $Arg^{84}$ from adjacent loops. Collectively, these interactions cause the subunit barrels to flatten slightly and become more tightly wrapped. Because the subunit barrels are part of a more extended β sheet that forms the hydrogen-bonded dimer (Figs. 1 and 2), and the barrel exteriors pack at the dimer-dimer interface (Figs. 2 and 4), changes in barrel curvature effect both hydrogen-bonded dimer geometry and dimer-dimer packing. The net result of the change in subunit barrel curvature is to alter the twist of β sheet that connects dimer subunits, which produces a slight increase in the angle between the barrel domains. The tetramer adjusts to the changes in dimer sheet twist and preserves the complementary sheet packing by a 5.4° rotation of the dimer subunits around the corresponding tetramer dyad axis (Fig. 4).

The unusually high affinity of streptavidin for biotin reflects participation of a number of factors, the analogs of which have been previously encountered individually in other protein-ligand interactions. These factors include oriented dipole arrays to stabilize bound oxyanions [for example, the oxyanion hole in serine proteases (*24*)], hydrogen-bond dipole networks to alter charge distribution on bound ligands [such as the serine protease charge relay system (*24*)], and disorder-order transitions to sequester bound ligands from the solvent environment [as in triose phosphate isomerase (*25*)]. In streptavidin, these factors, together with quaternary changes in structure, combine to produce both strong binding and a high activation energy for dissociation that characterize the near irreversibility of the biotin:streptavidin interaction.

## REFERENCES AND NOTES


1. L. Chaiet and F. J. Wolf, *Arch. Biochem. Biophys.* **106**, 1 (1964).
2. N. M. Green, *Adv. Protein Chem.* **29**, 85 (1975).
3. P. R. Langer, A. A. Waldrop, D. C. Ward, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6633 (1981); J. J. Leary, D. J. Brigati, D. C. Ward, *ibid.* **80**, 4045 (1983).
4. M. J. Cox and P. C. Weber, *J. Appl. Cryst.* **20**, 366 (1987); *J. Cryst. Growth* **90**, 318 (1988).
5. P. C. Weber, M. J. Cox, F. R. Salemme, D. H. Ohlendorf, *J. Biol. Chem.* **262**, 12728 (1987).
6. A. Pahlet *et al.*, *ibid.*, p. 13933.
7. A. J. Howard *et al.*, *J. Appl. Cryst.* **20**, 383 (1987).
8. J. Wolff and I. Covelli, *Eur. J. Biochem.* **9**, 371 (1969).
9. T. C. Terwilliger *et al.*, *Acta Cryst.* **A43**, 1 (1987).
10. T. C. Terwilliger and D. Eisenberg, *ibid.* **A39**, 813 (1983).
11. The agreement between multiple observed diffraction intensities, $R_m$ = {rms $[(I_j - \langle I \rangle)/\sigma_j]$} (rms, root-mean-square), for 84,062 observations of 9,225 unique reflections to 2.6 Å resolution was 0.054 for the streptavidin:biotin complex, and the $R_m$ for apostreptavidin native crystals was 0.052 for 89,393 observations of 13,496 observations to 1.7 Å resolution. The fractional difference in diffraction amplitudes to 2.2 Å resolution between native and derivative crystals of apostreptavidin was 0.14 for iodinated apostreptavidin and 0.19 for the crystal soaked in $K_2Pt(SCN)_6$. In the apostreptavidin structure determination by multiple isomorphous replacement techniques (*10*), the mean figure of merit for 4640 reflections to 2.6 Å resolution was 0.62, and the ratio of the mean heavy atom scattering factor to the mean lack of closure ($f_H/E$) was 1.4 for acentric reflections of the platinum derivative and 0.8 for those of the iodinated apostreptavidin.
12. T. A. Jones, *J. Appl. Cryst.* **11**, 268 (1978).
13. —— and S. Thirup, *EMBO J.* **5**, 819 (1986); B. C. Finzel, *Proteins Struct. Funct. Genet.*, in press.
14. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
15. J. W. Ponder and F. M. Richards, *ibid.* **193**, 775 (1987).
16. W. A. Hendrickson and J. H. Konnert, in *Biomolecular Structure, Function, Conformation and Evolution*, R. Srinivasan, Ed. (Pergamon, Oxford, 1980), vol. 1, pp. 43–57.
17. M. Fujinaga and R. Read, *J. Appl. Cryst.* **20**, 517 (1987).
18. The apostreptavidin tetramer was positioned with a twofold symmetry axis coincident with a crystallographic dyad axis of the streptavidin:biotin unit cell. The correlation coefficient between observed and calculated structure factors was calculated for each orientation of the tetramer as it was rotated 180° in 3° steps about, and translated in 1 Å increments along, the crystal twofold axes.
19. A. T. Brünger, *J. Mol. Biol.* **203**, 803 (1988).
20. P. Weiner and P. A. Kollman, *J. Comput. Chem.* **2**, 287 (1981); S. J. Weiner *et al.*, *J. Am. Chem. Soc.* **106**, 765 (1984).
21. B. C. Finzel, *J. Appl. Cryst.* **20**, 53 (1987).
22. F. R. Salemme, *Prog. Biophys. Mol. Biol.* **42**, 95 (1983).
23. C. E. Argarana, I. D. Kuntz, S. Birken, R. Axel, C. R. Cantor, *Nucleic Acids Res.* **14**, 1871 (1986).
24. J. Kraut, *Annu. Rev. Biochem.* **46**, 331 (1977).
25. T. C. Alber *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 603 (1987).
26. The structure of a streptavidin:biotin complex has been determined independently by W. A. Hendrickson *et al.* (*Proc. Natl. Acad. Sci. U.S.A.*, in press).


4 November 1988; accepted 12 December 1988